D3

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 7/00, C12P 21/06
C12Q 1/66

A1

(11) International Publicati n Number: WO 93/16172

(43) International Publication Date: 19 August 1993 (19.08.93)

(21) International Application Number: PCT/US93/00913

(22) International Filing Date: 2 February 1993 (02.02.93)

(30) Priority data:
07/833,431 7 February 1992 (07.02.92) US

(71) Applicants: ALBERT EINSTEIN COLLEGE OF MEDICINE OF YESHIVA UNIVERSITY [US/US]; Yeshiva University, 1300 Morris Park Avenue, Bronx, NY 10461 (US). UNIVERSITY OF PITTSBURGH [US/US]; Cathedral of Learning, Pittsburg, PA 15260 (US).

(72) Inventors: JACOBS, William, R., Jr. ; 163 Fordham Street, City Island, NY 10464 (US). BLOOM, Barry, R. ; 61 Summit Drive, Hastings-on-Hudson, NY 10706 (US). HATFULL, Graham, F. ; 6 Forbes Terrace, Pittsburgh, PA 15217 (US).

(74) Agent: GEORGE, Kenneth, P.; Amster, Rothstein & Ebenstein, 90 Park Avenue, New York, NY 10016 (US).

(81) Designated States: AU, BB, BG, BR, CA, FI, HU, JP, KR, LK, MG, MW, NO, PL, RO, RU, SD, SE, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG).

Published
With international search report.

(54) Title: MYCOBACTERIAL SPECIES-SPECIFIC REPORTER MYCOBACTERIOPHAGES

(57) Abstract

This invention relates to mycobacterial species-specific reporter mycobacteriophages (reporter mycobacteriophages), methods of producing such reporter mycobacteriophages and the use of such reporter mycobacteriophages for the rapid diagnosis f mycobacterial infection and the assessment of drug susceptibilities of mycobacterial strains in clinical samples. In particular, this invention is directed to the production and use of luciferase reporter mycobacteriophages to diagnose tuberculosis. The mycobacterial species-specific reporter mycobacteriophages comprise mycobacterial species-specific mycobacteriophages which contain reporter genes and transcriptional promoters therein. When the reporter mycobacteriophages are incubated with clinical samples which may contain the mycobacteria of interest, the gene product of the reporter genes will be expressed if the sample contains the mycobacteria of interest, thereby diagnosing mycobacterial infection.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SK	Slovak Republic
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
UA	Ukraine
US	United States of America
VN	Viet Nam

MYCOBACTERIAL SPECIES-SPECIFIC REPORTER MYCOBACTERIOPHAGES

STATEMENT OF GOVERNMENT INTEREST

This invention was made with government support under NIH Grant Number AI26170.

FIELD OF THE INVENTION

This invention relates to mycobacterial species-specific reporter mycobacteriophages (reporter mycobacteriophages), methods of making such reportermycobacteriophages, and the use of such reporter mycobacteriophages, for example, to rapidly diagnose mycobacterial infection and to assess drug susceptibilities of mycobacterial strains in clinical samples. Specifically, this invention relates to the use of mycobacterial species-specific luciferase reporter mycobacteriophages to diagnose tuberculosis and to assess the drug susceptibilities of the various strains of Mycobacterium tuberculosis (M. tuberculosis).

To produce the mycobacterial species-specific reporter mycobacteriophages of the invention, transcriptional promoters and reporter genes are introduced into the genomes of mycobacterial

species-specific mycobacteriophages. These reporter genes may be the genes for luciferase or the ß-galactosidase gene, and provide the DNA which encodes production of a gene product. The reporter mycobacteriophages may be used by incubating same with samples which may contain the specific mycobacteria of interest. If the mycobacteria of interest is present, then the reporter mycobacteriophages introduce the recombinant nucleic acids which encode expression of the gene product into the mycobacteria of interest, and the mycobacteria then express the gene product. The expressed reporter gene product may be detected by a suitable assay, for example, through the detection of photons or the conversion of an easily assayable chemical reaction. The presence of such gene product indicates that the sample contains the mycobacteria of interest, and hence the mycobacterial species-specific reporter mycobacteriophages may be used to detect and thereby diagnose the specific mycobacterial infection. In addition, since signals may not be generated by cells which are not metabolically active in the presence of antibiotics, the mycobacteria species-specific reporter mycobacteriophages of this invention may be used to assess the drug susceptibilities of various strains of mycobacteria. If antibiotic drugs are added to the sample containing the reporter mycobacteriophages and the gene product

is detected, the mycobacteria is metabolically active and hence resistant to the antibiotic drug.

BACKGROUND OF THE INVENTION

In 1990, there was a 10% increase in the incidence of tuberculosis in the United States. In addition, there has been an increase in the appearance of clinical isolates of tuberculosis that are resistant to antibiotics used to treat the disease. This problem is exacerbated by the length of time that is currently needed both to diagnose tuberculosis, and to determine the drug susceptibilities of various strains of M. tuberculosis. As a result, patients with M. tuberculosis may remain infectious for long periods of time without being treated, or may be treated with a drug to which the bacterial strain is resistant. Therefore, a need has arisen in the field for a method of diagnosis of M. tuberculosis (and other mycobacterial infections) which is rapid, sensitive and specific, which method is also capable of assessing the drug susceptibilities of the various strains of M. tuberculosis and other mycobacterial strains. It is critical that a mycobacterial strain be assessed for drug resistance rapidly because a patient infected with a strain of M. tuberculosis or another mycobacteria must be treated immediately with the particular antibiotic drug(s) to which the strain is not resistant, and not with antibiotic drug(s) to

which the strain is resistant, or the patient may die.

Currently, the most rapid test available for the diagnosis of M. tuberculosis is the staining of sputum samples for acid-fast bacilli, which is a tedious procedure, and which procedure has high sensitivity. Alternative methods for diagnosis require cultivation of the bacilli for approximately two to six weeks followed by classification of the cultured organism. Typical diagnostic tools include biochemical tests, analysis of mycolic acids and serotyping. All of these tests are time-consuming. More recently, the use of oligonucleotide probes and Polymerase Chain Reaction have been suggested for the identification of M. tuberculosis species. Although these methods may be useful approaches, their uses in a clinical setting have not yet been determined. Further, these methods do not distinguish between live and dead organisms, and are therefore of limited use in the determination of drug sensitivities of clinical isolates.

In addition, Mycobacterium avium (M. avium) is a mycobacteria which is often found in immunosuppressed patients. This mycobacteria is typically disseminated throughout the bodies of immunosuppressed patients, such as AIDS patients, and causes M. avium infection. Because this mycobacteria often causes death in immunosuppressed patients, it is

necessary to be able to diagnose and assess the drug susceptibilities of the various strains of M. avium.

It is therefore an object of this invention to construct broad mycobacterial host range and mycobacterial species-specific reporter mycobacteriophages.

It is another object of this invention to provide mycobacterial species-specific reporter mycobacteriophages which may be used to rapidly diagnose mycobacterial infections.

It is still another object of this invention to provide mycobacterial species-specific reporter mycobacteriophages which may be used to rapidly assess the drug susceptibilities of different strains of mycobacteria in clinical samples.

It is yet another object of this invention to provide mycobacterial species-specific reporter mycobacteriophages wherein the reporter genes are luciferase genes, which mycobacterial species-specific reporter mycobacteriophages may be used to rapidly diagnose mycobacterial infections and to rapidly assess the drug susceptibilities of various strains of mycobacteria.

It is a further object of this invention to provide mycobacterial species-specific luciferase gene reporter mycobacteriophages which may be used to rapidly diagnos tuberculosis and assess the drug

susceptibilities of the various strains of M. tuberculosis.

SUMMARY OF THE INVENTION

This invention relates to broad host range and mycobacterial species-specific reporter mycobacteriophages, (reporter mycobacteriophages), methods of producing such reporter mycobacteriophages, and the use of such reporter mycobacteriophages to rapidly diagnose mycobacterial infection, such as M. tuberculosis, and to distinguish which strains of the mycobacteria are drug-resistant. To produce these reporter mycobacteriophages, reporter genes and transcriptional promoters are introduced into the genomes of mycobacterial species-specific mycobacteriophages. The promoter and reporter gene-containing mycobacteriophages (reporter mycobacteriophages) are then incubated with a clinical sample which may contain the mycobacteria of interest, such as M. tuberculosis. The reporter mycobacteriophages are specific for the mycobacteria which is sought to be detected. The reporter mycobacteriophages efficiently introduce the recombinant nucleic acids which encode the expression of the reporter gene's gene product into the mycobacteria of interest, and the mycobacteria then express the gene product. A substrate or other means

capable of allowing for the detection of the gene product is then added to the sample. If the gene product or the signal generated by the gene product is detected, the presence of the infectious mycobacteria is known, thereby diagnosing the disease. To assess drug susceptibility of mycobacteria, drugs such as antibiotics ma; be added to a sample containing the reporter mycobacteriophages of this invention. If the mycobacteria are susceptible to a drug after exposure to the drug, the mycobacteria will be killed. However, drug-resistant mycobacteria will continue to be metabolically active in the presence of the drug, and will continue to express the detectable gene product of the reporter genes.

The preferred reporter genes of the present invention are the Firefly luciferase lux gene (FFlux), the luciferase lux genes of Vibrio fischeri, the luciferase lux genes of Xenorhabdus luminescens and the E. coli ß-galactosidase gene (lacZ). The preferred promoters of the present invention are hsp60 and L5 gene 62 promoter, and the preferred mycobacteriophages are L5, TM4 and D56A. These reporter mycobacteriophages are preferably used for the rapid diagnosis of tuberculosis and M. avium infection, and the accurate assessment of drug susc ptibilities of the various strains of M. tuberculosis and M. avium.

BRIEF DESCRIPTION OF THE DRAWINGS

The above brief description, as well as further objects and features of the present invention, will be more fully understood by reference to the following detailed description of the presently preferred, albeit illustrative, embodiment of the present invention when taken in conjunction with the accompanying drawings wherein:

FIGURE 1 represents the genome organization of mycobacteriophage L5;

FIGURE 2 represents a luciferase shuttle plasmid pYUB180 wherein reporter gene FFlux is fused to the BCG hsp60 promoter;

FIGURE 3 represents the amount of luciferase activity of M. smegmatis which contains the pYUB180 shuttle plasmid and the FFlux gene;

FIGURE 4 represents the effect of various antibiotic drugs on the metabolic activity of control mycobacteria and drug resistant mycobacteria in the presence of reporter mycobacteriophages which contain luciferase reporter genes;

FIGURE 5 represents shuttle plasmid phAE39 wherein the reported gene is FFlux, the promoter is hsp60, the phage is TM4 and the cosmid is pYUB216.

FIGURE 6 represents luciferase activity of M. smegmatis cells infected with shuttle phasmids phAE39; and

FIGURE 7 represents a flow chart for cloning different promoters into TM4::lux shuttle phasmid phAE39.

DETAILED DESCRIPTION OF THE INVENTION

This invention is directed to mycobacterial species-specific reporter mycobacteriophages, (reporter mycobacteriophages), methods of producing such reporter mycobacteriophages and the use of such reporter mycobacteriophages for the rapid diagnosis of mycobacterial infections and the accurate assessment of mycobacterial drug susceptibilities.

In order to produce such reporter mycobacteriophages, mycobacterial species-specific mycobacteriophage genomes are modified by introducing therein transcriptional promoters and reporter genes whose gene product can be sensitively detected. The reporter mycobacteriophages may then be incubated with clinical samples suspected of containing the mycobacteria of interest, either directly of after culture, and the samples tested for the presence of the reporter gene product, thereby diagnosing mycobacterial infection.

The method of this invention allows for rapid diagnosis because only the amount of time necessary for the reporter mycobacteriophages to infect their host cells and the amount of time necessary for the host cells to synthesize the reporter gene product are

required to allow for diagnosis. Typically, the amount of time required for the reporter mycobacteriophages to infect their host cells and for the host cells to synthesize the reporter gene product is between ten minutes and sixteen hours.

The assessment of drug susceptibilities with the reporter mycobacteriophages of this invention is accurate because the reporter mycobacteriophages only allow for the detection of metabolically active mycobacterial organisms, the presence of which metabolic activity indicates that a drug has not killed the mycobacteria and that the mycobacteria is resistant to the drug.

To enhance diagnosis specificity, a series of similar reporter mycobacteriophages, each of which having well-defined but different specificities for mycobacterial species, is selected.

Mycobacteriophage L5, a temperate virus with a broad host-range among mycobacteria, is the most thoroughly characterized of the mycobacteriophages. L5 particles are morphologically similar to the family of phages that includes phage g and contain a linear dsDNA genome with cohesive ends. The inventors have determined the DNA sequence of the entire gene as well as several gene functions. The DNA sequence of the L5 mycobacteriophage is as follows:

* * * S E Q U E N C E * * *

```
   1  GGTCGGTTAT GCGGCCGAGC CATCCTGTAC GGGTTTCCAA GTCGATCAGA GGTAGGGGCC
  61  GGCACAGAAA CCACTCACAT CAGGGCTGTG CGCCTCCAGG GCGCGTGAAC TCCCACACCC
 121  CGGTGTAGTT ACATCCCGGA ATTGTCTCAG CGCCTCTCAG GGCGCTTCTC ATAAACAGTG
 181  ATCTACGCCA CTCCTGACGG GTGGCTGTCA AGGATACTCA CCTTCCCTAC TAATGAGGGG
 241  CTAAGAGCCC CTCTCTATAG AGCGCCGCAC AGGCGGCGCG ATAAGAGCGC CACCAGGCGC
 301  TCATCTAAAG ACCGGCCTTG AAGGGCCGGT CATAGAGATC TATTCGATCC GGCAACCGCC
 361  GGATCTCAAG GCCGCGCCAG TGCGCGGCCC TATAGAGGGG TGACTCAACT GTGCATGGCA
 421  CTCGCTCGAG TGCCCACTGG AGCACTCAAC CGGGGAAGTT CGACGTTCTC AACCTGCGAA
 481  TGACGTTTGA ATCGTCATCC GCGTACGAAA TCCCCGATCT GCGGCCGACC GACTTCGTGC
 541  CGGCCTATCT CGCGGCCTGG AATATGCCGC GTCACCGCGA TTACGCCGCC AAGAACGGCG
 601  GCGCGCTGCA CTTCTTCCTT GACGATTACC GGTTTGAGAC CGCGTGGTCG TCCCCCGAGC
 661  GCCTTCTCGA CCGCGTAAAG CAGGTCGGCG CTGCACTCAC GCCGGATTTC AGCCTCTGGA
 721  CGAACATGCC GAAGGCGGCG CAGCTATGGA ACGTCTACCG CTCCCGCTGG TGTGGCGCGT
 781  ATTGGCAGTC GGAAGGAATC GAGGTGATTC CGACGGCGTG TTGGGCGACT CCCGACACGT
 841  TCGATTTCTG TTTCGACGGG ATCCCGATGG GATCGACCGT CGCAATTTCT TCGATGGGCA
 901  TTCGCTCTTC AAAAGTCGAC CAGGAGCTTT TCCGGTACGG ACTACGCGAA CTCATCGATC
 961  GCACTCAACC GCAACTGCTT TTGGCATATG GCCAGCTTCG GCATTGCGAC GACATGGATT
1021  TACCAGAGGT CCGCGAATAC CCGACCTACT GGGACAGACG ACGAAAGTGG GTAACTGCCG
1081  ATGGGAGGCC GGGGAAGTAA AGGCGGCCCC GGTCCCGGAA CCGGAGCACG CAACCGCAGA
1141  GGCGCTGGAG CCCCCGGATC GGGCGGCGTA GGCGGCGTCG GAGGCGGGGG TGGAGCTGCA
1201  GGGAGCAGCG GAGGCGGCAA GGGAACGGCA GCGCCGGTAC CGGAGGCGTC ACCGGTGGCG
1261  GCGGAAGTGG AGCCGGCGGC GGTGGCAGCA GCCCCAACAC CCCGGTGCCC CCCACCGAGC
1321  TGGAGAAGAA GCGCGGCGAA TACAACCAGA TCGCCATCGA CGCCCAGAAA CAGCACGCGC
1381  CCACCGATGA GAAGCGCGAG GCCAAGCGCA AGCAACTGAT GGATCGAGTC GGAGGAGACT
1441  GGCAGGCTTT GGACCCGGAT CACCACGACG CCATCAAGGT GGCGATGGAT GACGCCATGC
1501  GGAAGATCCT CTCCGAGGAG GAGATCGTCC ACCGCACCAA GCACTTCGGC GACCTACTCG
1561  ACTCCGGTCG ACTCAAGTCG CTGTTCGAGG TCGGCTTCTC AGCCGGTGGC GACACCCCGA
1621  CCGAACGCGC CCTCCTCGAG GACGCCTGGT TCGGCGCAGG CAAGGTTCCC CCGATCTACT
1681  CGGCAATCGA GTTCAACGGC GCTCCGACAG CCGGCCTCGG CATGTACGGC GGCACCAAGC
1741  TCTACATGAA GGACTCGGTC AAGGACCGCG TCACCGTGAC CATCGGCGAC TCGCTGATGT
1801  CGAGCTGGGA CGTATTCCCC GGCCGTCCTG GCGACGGCGT GGGGCTGTGG GCCAGCCTGT
1861  CGAAGATCGA GGGGCTGGTC GATCCGAGCA AGACCCGCGA AGAGAACATG CAGGCGGTGT
1921  ACGACTCGTT CAAGAAGTAC GGCACCCTGG ACGGCTTCAT CGAGGCGCAG ATCCACGGCG
1981  GCGTCCTGGT CGAGGACATC AAGAAGGTCG TGTTCACGCA GCCGCCGAGC CCGATCTTCA
2041  CCGATAAACT GGACGAACTT GGAATCCCGT GGGAGGTGCA GTAATGGCGC AGATGCAGGC
2101  GACACACACA ATCGAGGGGT TCCTGGCTGT CGAGGTGGCC CCTCGGGCGT TCGTCGCAGA
2161  GAACGGCCAC GTACTGACCC GGCTGTCGGC CACGAAGTGG GGCGGTGGCG AGGGTCTCGA
2221  GATCCTCAAC TACGAGGGTC CAGGGACCGT CGAGGTCTCC GACGAGAAGC TCGCCGAAGC
2281  CCAGCGGGCC AGCGAGGTCG AGGCTGAACT TCGCCGCGAG GTCGGCAAGG AGTGAGCTGG
2341  GCCGGCTCAG GCCGGCGACA GGAACTACCA GAGGACTGGG AGCTGAATTA CCGGCTCCCG
2401  GTCCTTTCTG CTGCCAACTG GCTTTGCCAG ATCAACGGTC CCGGATGCGT AAGGGCCGCA
2461  ACCGATGTCG ACCACATCAA GCGCGGGAAC GACCACAGCC GGTCCAATCT GCAGGCAGCC
2521  TGCCATGTCT GTCACGGCAA GAAATCAGCC GCCGAGGGCG TAGCCCGACG GCGGGAACTT
2581  AGAGCCCGGA GGAAGCGACC ACCCGAACGC CATCCTGGGC GTCGATAAGC GGGCCAGGTG
2641  CCCGCTCCAC CCAGGAGGTG AACAGTGGGC ACGCGAGGCC CAATCGGAAA ACGAGATGAA
2701  GAGCGGGTTC GTCGGAACAC CCCGGACAGT CCAACCGACA CGATCCAGAT GCCCGGTCTG
2761  GTGACGATCC CCGAGATGGG CGATCTAAGC CACGACCGCC GCACGCACCA GCTCGTCAAG

2821  GACATGTACG AGTCGATCAA GCAGTCGGCA GCCGTGAAGT ACTACGAGCC GACCGACTGG
2881  CAGATGGCCC GACTCGCCCT CTACACACTT AACCAGGAAC TCATCGCAGC CGAGAACAAC
2941  GGCAAGCCCG TGGGCGCGAT GAAGCTCACT GCCATCAACC AGATGCTCTC CGCGCTGCTG
3001  CTGACCGAAG GTGACCGACG CCGCGTCCGA CTCGAAGTCG AACGAGCACC CGCTGACCCG
3061  ACAGGCGGGA AGGTCGTTGA CGTGACCGAC GTGCTCAAGC AGCGCCTCGC CAAGGCGAGC
3121  GGCGGGAGCT GATGGTCCCC CGAGGGGTTT CTAGAGCCGC TGCCGCTACC AGCCGCTCCC
3181  CCTCGGGGTA GACATCGAAA GGAACCACAT GGCCGACCTC GGCAACCCAC TCGACCTCGA
```

```
3241 GATGCTCTGC CTGGTCACAG GCCGGGACTT CCGCTGGACC ATCGATTACC CGTGGGGTCC
3301 GGGAGAGCTG TTCCTCGAAC TCGAGACCGG CGGCGAACAC AACGCGCTGC ATCAGGTCTA
3361 TGTCACCGGG GCGACCGGAG GCACGTACAC GCTGAACGTC AACGGCACCA ACACCCCGGC
3421 CATCGACTAC AACGACGTGT CGGAGAATCC GCAGGGGCTG GCAGGCGACA TCCAAGACGC
3481 TCTGGACGCA GCCGTCGGAG CCGGAAACGC TGTCGTGCAT CCGGTCTCGC TGTTCCCTGC
3541 GTGGACACTG AACTTCAACC TCAACGCCAG CAAGCCGCTC ACCGAGCAGT TGGTCAACAC
3601 GATCAACAAG GCCGCGAACG ACTTCTTCGA CACGTTCGAC CAACTACTTG GGGTCGACGT
3661 GGAGATGACG GTCACCGACA CCCTGAACTT CAAGCTCAAG GTGACCTCGC GGCGCTCGTT
3721 CGATGAGGTC GGTGTCGTCA CGTTCGCGGT CGACGTGACC AGCCAGGCAG TCATCAACTT
3781 CTTCAACTCC GTCGCCGAAC TCACCGGAGC GGTGAACACC GTCAACGTCG ACTTCTACTG
3841 GAACCGGACG TATGACATCG AGTTCACCGG ATCCCTTGGG CTGCAGCCGA TTCCGGCTAC
3901 TACAGCCGAC ATCACCAACC TGGCGGGTAC CAGCAAGGCC GTCTCAGTCA CGGTGGTCGA
3961 GCCAGGAAAG AAGAGGCTGA CCATCTGGCC GTTCACGGTC AACGGTGAAA CCGCAACCAT
4021 CAAGGTCGAG TCCGAAGAGG CCGACAAGAT CCCCAACCGC TGCCGCTGGC AGTTGGTTCA
4661 CATGCCGACC GGCGAGGCAG CCGGCGGCGA TGCAAAGCAG CTCGGCCGCG TTTACCGACA
4141 GCCGAGGTAA CACCGCACCC ATCAGAGATG GTGGGCCAGA CGGCCTTCGG GCCGTCCCCT
4201 GACGTGTAGC TCAATGGCAG AGCGCCCGAC TGTTAATCGG GTGGTTGAAG GTTCGAGTCC
4261 TTCCATGTCA GCGAGGGCTG AACCGGACCC GTGTCCGGTG TAGGCACTTT CCGCAGGCGG
4321 TTCCCCAGAG CGTGGGGAGC CCCTGCCCTG TACACGTAGC TCAATTGGTA GAGCAGCGGT
4381 CTCCAAAGCC GCCGGTTCCA GGTTCGACTC CTGGCGTGTA TGCACACACC CCTGACTCCT
4441 GCTAGCGGAG TGTTCGCCTT TCGGGCCTGG GGTCTTTTTC CCCGTTCGTC TAATCGGTAA
4501 GACACCCGGC TCTGGACCGG GCAATTGAGG TTCGAGTCCT TGGCGGGGAG CCAACTTGAC
4561 ATCCACCCGA AAGGAACAAC ATGACCTTCA CAGTCACCCG CGAGAGAGCG CAGTGGGTCC
4621 ACGACATGGC CCGCGCTCGC GACGGTCTCC CCTACGCGTA CGGCGGGGCG TTCACCAACA
4481 ACCCGAGGGT GTCGACTGAC TGCTCTGGCC TGGTGCTGCA GACCGGGGCT TGGTATGGAG
4741 GTCGCACCGA CTGGGTCGGA AACCGTTACG GCTCAACCGA ATCGTTCCGG CTCGACCACA
4801 AGATCGTCTA CGACCTAGGG TTCAAGCGGA TGCCCCGAGG CGGGCCAGCG GCCTTGCCGA
4861 TCAAGCCGGT GATGCTCGTC GGGCTCCAGC ACGGAGGCGG CGGGGTCTAC TCGCACACCG
4921 CTTGCACGTT GATGACGATG GACCACCCCG GTGGCCCGGT CAAGATGTCC GACCGAGGCG
4981 TCGACTGGGA GTCCCACGGC AACCGCAACG GCGTAGGCGT CGAACTTTAC GAGGGCGCAC
5041 GGGCATGGAA CGACCCTCTG TTCCATGACT TTTGGTACCT GGACGCAGTC CTCGAAGACG
5101 AAGGAGACGA TGACGAATTG GCTGACCCAG TTCTAGGGAA GATGATCCGC GAGATCCACG
5221 CGTGCCTGTT CAATCAGACC GCGTCGACCA GCGATCTGGC GACCCCTGGT GAAGGCGCTA
5281 TCTGGCAGCT ACACCAGAAG ATCCACTCGA TTGACGGCAT GCTCCACCCG ATCCACGCTG
5341 AGCGGCGCGC TCGCGCAGGC GATCTCGGTG AGCTGCACCG AATCGTGTTG GCCGCGAAGG
5401 GCTTGGGCGT GAAGCGCGAC GAGGTGACCA AGCGGGTCTA CCAGAGCATC CTCGCCGACA
5461 TCGAGCGGGA CAACCCCGAA GTACTTCAGC GATACATCGC AGAAAGAGGT GGCCTATGAG
5521 CCCCAAGATC CGACAGACCA TCTACCTGCT CGGCACCGCC GCCCCGGCAC TGCTGGGCAT
5581 CGTCCTGATC TGGGGCGGGC TCGACGCTGA GTCGGCGGCT GACCTCGGTG ACATCATTGC
5641 GGGCGTCGTG TCGATACTAG TCTCCGGTGC GCCGGCCGTA GCGGCAGGCA CCGTACGCAG
5701 CCAGCGCAAG GACGGCACGT TGTCCACCAG CCCGGTGGAT CAGGTCACCA AGGGCGTCGA
5761 GCAGGTGCTC GCGGCCAGGC AGAGTGCCGA GGCTGAAGTC GCGAAGGTCA AGCAGGCGCT
5821 GGAGACCGCC GTCAGCGGTT CTCTCCCCCA GCTCGGCCCG CTGGCCACGC AGATCCTCAA
5881 CGTGGCTGAC GACACCGTCT GGCGTCCATG AGCAAGCCCT GGCTGTTCAC CGTCCACGGC

5941 ACAGGCCAGC CCGACCCGCT CGGGCCTGGT CTGCCTGCCG ATACCGCACG GGACGTACTT
6001 GACATCTACC GGTGGCAGCC CATCGGCAAC TACCCGGCAG CGGCGTTCCC GATGTGGCCG
6061 TCGGTCGAAA AGGGTGTCGC TGAGCTGATC CTGCAGATCG AGCTGAAGCT GGACGCAGAT
6121 CCGTACGCGG ACTTCGCGCT GGCCGGCTAC TCGCAGGGAG CCATCGTGGT GGGCCAGGTG
6181 CTCAAGCACC ACATCATCAA CCCGAGAGGT CGACTGCACC GGTTCCTGCA CCGGCTCAGG
6241 AAGGTCATCT TCTGGGGTAA TCCGATGCGG CAGAAGGGCT TTGCCCACAC CGACGAGTGG
6301 ATTCACCAGG TCGCTGCCTC GGACACGATG GGCATCCTCG AGGACCGACT GGAGAACCTC
6361 GAGCAGTACG GCTTTGAGGT CCGCGACTAC GCGCACGACG GCGACATGTA CGCCTCCATC
6421 AAGGAGGACG ACATGCACGA GTACGAGGTG GCCATTGGCC GAATCGTGAT GAGCGCTAGG
6481 CGATTCATCG GAGGTAAGGA CTCCGTCATC GCCCAGCTCA TCGAGCTTGG ACAGCGTCCG
6541 ATCTGGGAGG GAATCGCGAT GGCCAGAGCC ATCATCGACG CCCTCACGTT CTTCGCCAAG
6601 TCGACCCAAG GCCCGAGCTG GCCGCATTTG TACAACCGCT TCCCGGCGGT CGAGTTCCTA
6661 CGACGAATCT GAGAAAGGAG GCGGGGTGAG CCTCAACAAC CACCACCCGG AGCTTGCCCC
6721 GTCTCCCCCT CACATCATCG GCCCGTCCTG GCAGAAGACG GTCGATGGTG AGTGGTATCT
6781 GCCTGAGAAG ACCCTCGGCT GGGGAGTCCT GAAGTGGCTC TCCGAGTACG TGAATACCCC
     TGGCGGGCAT GACGATCCGA ACCGTCTGGC GACGTTGATC GCGCTCTCCG AGGCAGGTCT
```

SUBSTITUTE SHEET

```
6841 TCTCGACAAC GAGAACATGT TCATCCCCAC CGACGAGCAG GTACGCCTGG TCCTCTGGTG
6901 GTACGCAGTA GATGACCAGG GCCAGTACAT CTACCGCGAG GGCGTGATCC GCCGGCTCAA
6961 GGGCTGGGGC AAGGATCCGT TCACCGCCGC GCTCTGCTTG GCGGAACTCT GTGGCCCCGT
7021 AGCCTTTTCA CACTTCGACG CCGACGGTAA CCCGGTCGGC AAGCCGCGTT CAGCCGCGTG
7081 GATCACCGTC GCGGCCGTCA GCCAGGACCA GACGAAGAAC ACGTTCTCGC TGTTCCCGGT
7141 GATGATCAGC AAGAAGCTGA AGGCCGAGTA CGGCCTGGAC GTGAACCGCT TCATCATCTA
7201 CTCCGCAGCC GGTGGCCGTA TTGAGGCAGC GACCTCGAGC CCCGCGTCGA TGGAGGGTAA
7261 CCGCCCGACG TTCGTCGTCC AGAACGAGAC GCAGTGGTGG GGCCAAGGCC CCGACGGCAA
7321 GGTCAATGAA GGCCACGCGA TGGCAGAGGT CATCGAAGGC AACATGACCA AGGTCGAGGG
7381 CTCCCGCACC CTGTCGATCT GCAACGCCCA CATCCCCGGC ACCGAGACGG TCGCCGAGAA
7441 GGCATGGGAC GAGTACCAGA AGGTCCAGGC AGGCGACTCT GTCGACACCG GGATGATGTA
7501 CGACGCGCTG GAAGCGCCGG CCGACACCCC GGTCTCCGAG ATCCCCCCGC AGAAGGAGGA
7561 TCCCGAGGGA TTCGAGAAGG GCATCGAGAA GCTCCGCGAG GGCCTGCTCA TCGCCCGAGG
7621 CGACTCCACC TGGCTGCCGA TAGACGACAT CATCAAGTCG ATTCTGTCGA CCAAGAACCC
7681 GATCACCGAG TCGCGGCGCA AGTTCCTGAA TCAGGTAAAC GCCGCTGAGG ACTCGTGGCT
7741 CTCACCGCAG GAATGGAACC GGTGCCAGGT CGACCTGGCC AAGTACCTGG ATAAGCACGG
7801 CAGGGAGTTC GCTCCGCTGC AGCGCGGTGA CCGGATCACC CTCGGGTTCG ACGGGTCGAA
7861 GTCCAACGAC TGGACCGCGC TCGTCGGCTG CCGTGTCAGC GACGGCCTGC TGTTCGTCAT
7921 CGACATCTGG GATCCCCAGA AGTACGGCGG GGAGGTTCCC CGCGAAGACG TTGACGCCAA
7981 GGTCCATTCG GCGTTCGCCC ACTACGACGT GGTGGCGTTC CGCGCCGACG TGAAGGAGTT
8041 CGAGGCGTAC GTCGACCAGT GGGGCCGGAC CTACAAGAAG AAGCTCAAGG TCAACGCCAG
8101 CCCGAACAAC CCGGTGGCGT TCGACATGCG CGGACAGCAG AAGAGGTTCG CGTTCGACTG
8161 CGAGCGACTC GAGGACGCGG TCCTTGAGGG CGAGGTCTGG CACGACGGCA ATCCCGTTCT
8221 GCGCCAACAC GTTCTGAACG CCAAACGACA CCCAACGAAC TACGACGCCA TCGCGATTCG
8281 CAAGGTCACG AAGGACTCCA GCAAGAAAAT CGACGCTGCA GTCTGCGCTG TCCTCGCGTT
8341 CGGGGCGAGA CAGGACTACC TCATGAGCAA GAAGGCCCGT AGCGGCCGGG TGGTGATGGT
8401 TCGATGACAG CACCGCTCCC CGGTATGGAG GAGATCGAAG ACCCCGCAGT CGTACGAGAA
8461 GAGATGATCT CGGCCTTCGA GGATGCTTCC AAGGATCTCG CCAGCAACAC CAGCTACTAC
8521 GACGCTGAGC GCCGGCCAGA GGCCATCGGC GTCACCGTCC CGAGAGAGAT GCAGCAACTG
8581 CTGGCTCACG TCGGATACCC CAGGCTCTAC GTCGACTCAG TCGCCGAGCG CCAGGCCGTC
8641 GAGGGTTTCC GCCTCGGCGA TGCCGACGAG GCTGACGAAG AGCTGTGGCA GTGGTGGCAG
8701 GCCAACAACC TCGACATCGA GGCACCACTG GGCTACACCG ACGCTTACGT TCACGGCCGG
8761 TCGTTCATCA CGATCAGCAA GCCAGACCCG CAGCTCGACC TGGGTTGGGA TCAGAACGTC
8821 CCGATCATCC GCGTCGAGCC GCCCACCCGA ATGCACGCCG AGATCGACCC CCGGATCAAC
8881 CGGGTGTCCA AGGCCATCCG AGTCGCATAT GACAAGGAGG GCAACGAGAT TCAGGCTGCC
8941 ACGCTGTACA CGCCGATGGA GACCATCGGC TGGTTCCGCG CTGACGGTGA GTGGGCTGAG
9001 TGGTTCAACG TCCCGCACGG TCTGGGCGTC GTTCCCGTTG TGCCGCTTCC GAACCGGACC
9061 CGGCTCTCGG ACCTGTACGG CACCAGTGAG ATCACGCCCG AGCTTCGGTC GATGACCGAC
9121 GCGGCGGCGC GCATCCTCAT GTTGATGCAG GCGACCGCCG AGCTGATGGG TGTCCCCCAG
9181 CGCCTGATCT TCGGCATCAA GCCCGAAGAG ATCGGCGTCG ACTCCGAGAC CGGCCAGACG
9241 CTGTTCGATG CGTACCTGGC CCGGATCCTG GCGTTCGAGG ACGCTGAGGG CAAGATCCAG
9301 CAGTTCTCTG CAGCCGAGCT GGCCAACTTC ACCAACGCGC TCGATCAGAT CGCCAAACAG
9361 GTCGCTGCGT ACACGGGATT GCCTCCCCAG TACCTGAGTA CCGCCGCAGA CAATCCGGCC
9421 TCCGCTGAGG CGATCAGGGC CGCTGAGAGC CGACTCATCA AGAAGGTCGA GCGGAAGAAC
9481 CTGATGTTCG GCGGCGCATG GGAAGAGGCC ATGCGGATCG CCTACCGGAT CATGAAGGGC
9541 GGCGACGTTC CCCCGGACAT GCTCCGCATG GAGACCGTCT GGCGAGACCC GAGCACTCCC
9601 ACCTACGCGG CCAAGGCCGA CGCAGCCACG AAGCTGTACG GCAACGGCCA GGGTGTCATC
9661 CCGCGTGAAC GTGCTCGCAT CGACATGGGC TACTCCGTCA AGGAGCGCGA AGAGATGCGC
9721 CGATGGGACG AGGAAGAGGC CGCAATGGGT CTCGGCCTGT TGGGCACGAT GGTCGACGCC
9781 GACCCGACGG TCCCAGGCTC CCCGAGCCCC ACGGCACCGC CGAAGCCACA GCCGGCCATC
9841 GAGTCGTCTG GTGGTGATGC GTGACCGCAG AGGAGTACGC GGCGGCTCAA GCCGCGATCA
9901 CTGCGGGTCT TGCCACATAC GTCCAGAGGT TCGCTTCGCT CTTCGTCGGT CCAGCTCTCG
9961 CTGTAGGTGA GTGGCTGCGA CTGCTGCAGG TGCTGTTCCC CGAAATCCAA CGGCGGTATG
10021 CAGATGCTGC CGCCTTGGGC AGGGACTTCT ACGACTCCCA ACGCGCACTA CACCACCCAG
10081 AGCTGCCCCG GAACGAGAGG TTCCGGGGAG AGCTTCGGTG GGAGTGGTTC GTCCAGAACA
10141 TGGAGCCCGC TCGAAAAGAG ATGTCGCAGG CCGACTCTCC GCCGAGTGCG ACCTCTAAGT
10201 TGGCTCTGGC CGCAGTTCGC GAAGTGGAGA TGGCAGCACG CCGACAGATC ATCGGCGCTG
10261 TCAAGAACGA TCCGGCCCCG CAGATCGTGC AGGGCTGGGC GAGGGTCGCC ACCGGGCGCG
10321 AAACATGCGC CTGGTGTCTG ATGCTCATCT CACGGGGTGC CGAGCTGAAT CACAAGGGCA
10381 ACTTCGCCTA CAGCTCAGCG GAAGCCGCAG GGCTCAACCT CGATGACGAG ACCGTGATCG
```

```
10441 ACCTCTGGAA CGAGTCCGGT CACGACCTTG AGAAGTTCCG CGAGGAGACC AGAGAGGACT
10501 TCGAGAAGTG GCACGCAGGG TGCGACTGTC TGGTGGTCCC GGTCTTCGAT GTGCAGAACT
10561 GGCCCGGAAG AGACGCTGCC CTACGGGCGC AGCAACTTTG GATCGAAGCC AGCGACGAAG
10621 CTGACGACCT CATTGCGTCA GGCAAGGCCC GCTCCAAGAA CAAGAACACG GAGACGCTCA
10681 ACGCGCTCCG ACGCCGCCTA GCACGCGGCG AAATCACCAT GTCCAACTAC GCCCTCGCTG
10741 CGTAGTCCCT CGAACCCCAG GTGGGTTCTC TCAACATGCC CAGGAGGCGA AAACACATGT
10801 CCGACAACCC CACTCCCGAG AGCACCCCAG AGGCCGAGAC CCCGGAGGTC GAGAAGCCGA
10861 TGGAACCGCA GGGCAAGGTC TTCGATGAAG CGTACGTTCA GTCGCTTCGC CAGGAGGCTG
10921 CAGCCGCTCG GGTGGCGAAG AAGGACGCCG TAGAAGCGGC AGAGGCTCGA GTGAAGGCCG
10981 AGTACGAGGC CAAGCTCGCT GAGCGCGACA CCGCTTACAC CGAACTGCAG AACCAGTTGG
11041 GACAGGCGTG GATTGAGCTG GAGAAGGTCT ACCTCTCTCT CGACGCCAAG GTGCCCAACG
11101 ACAAGGTTCG GGCGTTTGTC GAGATCCTCG AAGGCAACGA CAGGGACAGC ATCGCTGAGT
11161 CAGTGAAGTC CCGTCTGGAG CTGGTCGGCG GATTCGGCAA CAAGACCCCG AGTCCTGCGT
11221 TCGACCCGTC TCAGGGTCGC GGCGGTAAGC CGCCGATCCC GCTGAACGGT GACCCGATCC
11281 TCGAGGCCAT CAAGGCCGCT GTCGGGATCA AGAAGTAACC CACCCAACAG ATCTCAAGGA
11341 GAGATAAACA ATGGCAGTCA ACCCTGACCG CACCACGCCG TTCCTCGGCG TGAACGACCC
11401 CAAGGTCGCG CAGACCGGCG ACTCGATGTT CGAGGGCTAC CTCGAGCCCG AGCAGGCCCA
11461 GGACTACTTC GCCGAAGCGG AGAAGATCTC CATCGTCCAG CAGTTCGCCC AGAAGATCCC
11521 GATGGGCACG ACCGGCCAGA AGATCCCGCA CTGGACCGGC GACGTGAGTG CGTCGTGGAT
11581 CGGTGAAGGC GACATGAAGC CCATCACCAA GGGCAACATG ACCTCGCAGA CCATCGCCCC
11641 CCACAAGATC GCGACGATCT TCGTGGCCTC GGCGGAAACC GTCCGTGCGA ACCCGGCCAA
11701 CTACCTGGGC ACCATGCGGA CCAAGGTCGC GACCGCCTTC GCGATGGCGT TCGACAACGC
11761 CGCGATCAAC GGCACCGACA GCCCGTTCCC GACCTTCCTA GCGCAGACCA CCAAGGAGGT
11821 CTCGCTGGTG GACCCGGACG GCACCGGCTC CAACGCCGAC CTCACCGTCT ACGACGCGGT
11881 CGCCGTCAAC GCCCTGTCGC TGTTGGTCAA TGCCGGCAAG AAGTGGACCC ACACTCTGCT
11941 GGACGACATC ACCGAGCCGA TCCTCAACGG CGCGAAGGAC AAGAGCGGTC GCCCGCTGTT
12001 CATCGAGTCG ACCTACACCG AGGAGAACAG CCCGTTCCGC CTCGGTCGGA TTGTGGCCCG
12061 TCCGACCATC CTGAGCGACC ACGTCGCCTC GGGCACGGTC GTCGGCTACC AGGGTGACTT
12121 CCGCCAGCTC GTCTGGGGCC AGGTCGGCGG CCTGTCCTTC GACGTGACGG ATCAGGCGAC
12181 TCTGAACCTG GGCACCCCCC AGGCTCCGAA CTTCGTCTCG CTGTGGCAGC ACAACCTCGT
12241 CGCAGTCCGA GTCGAGGCCG AGTACGCCTT CCACTGCAAC GACAAGGACG CGTTCGTCAA
12301 GCTCACGAAC GTGGACGCCA CCGAAGCCTG ATCCAGGCTT GACATCCACC GGGAGGGGGC
12361 TCCTTCGGGA GCCCTCTCCT GATGTGGAGC AGGAAGGACC ACATGCGAAT CCAGTCCACC
12421 CTCAACGGCG GTTTCGCCGA GGTTTCCGAG GAGTTCGCCA AGCAGTTGAT CGCCACTGGC
12481 GGCTGGAAGG TGCCCCGGAA ACCGCGCAAC ACCAAGACCA AGACCGCTCC TGAGGAGCCC
12541 AAGAACGAGG AGTAACCCGT GGCCTACGCG ACCGCCGAAG ACGTTGTGAC GTTGTGGGCC
12601 AAGGAGCCTG AGCCCGAAGT GATGGCGCTG ATCGAGCGCC GGCTCCAGCA GATCGAGCGC
12661 ATGATCAAGC GCCGGATCCC CGACCTGGAC GTGAAAGCCG CTGCGTCGGC GACGTTCCGG
12721 GCCGATCTGA TCGACATCGA AGCTGATGCT GTTCTGCGCC TCGTGCGTAA CCCGGAGGGC
12781 TACCTCTCGG AGACCGACGG TGCGTACACC TATCAGCTCC AGGCCGACCT GTCGCAAGGC
12841 AAGCTCACCA TCCTCGATGA GGAGTGGGAG ATCCTCGGGG TCAACTCCCA GAAGCGCATG
12901 GCGGTCATCG TCCCGAACGT GGTGATGCCG ACGTGAGCGC GAGCGACCGA CACCGCGCCC
12961 CGATTGTCTA TCCGCCTGGC ACTCAGGCGG TTACGCCGGA TCGGGTCAAC GCGTTTGACT
13021 GCGATCACGA AGCTGATCCT CCGGTGTGCC GGTGCGTCCA CGACTGGCGC ATCGAGTGGG
13081 GAAACGTCAA GAAGGCCACC GCCAGATCAC GGTCGGCGGT GCTCTGATGA GCCTCCTCGA
13141 CACCGGTGCC CGGTACCAGA CCTGCATCGT CTACCCCGAA GAGATGGTCA TCGACTCCGA
13201 TGGCAACAAG CGGACCAGGC CGTCGAATAC CGGCATCCCG GCCATCGCAC GGTTCCAGGT
13261 AGCCAACCAG TCTGGTACGT CGGCACGACG TGCTGAGCAG GACAACGAGG GGTTCGAGAC
13321 CGAGAAGGTC TACCGGATGC GGTTTCCCCG CTCGTTCACC AAGGAGCACG GCATCCTCGG
13381 GGCCCAGTCC CAGATCGAGT GGCGAGACCA GCGGTGGGCG CTCTTCGGAG ACGCCACCGT
13441 CTACGACTCA TCCCCTGCGT TGGCGCGGGT CGACTACACG ATCAAGAGGT ACTGATGGCC
13501 AAGGTCTACG CGAACGCGAA CAAGGTCGCG GCCCGGTACG TCGAGACGAG GGACGCCGTC
13561 CGAGACGAGC GGAACAAGGT CACCCGTCGA GCCAAAGCCA ATCTGGCGCG GCAGAACTCG
13621 ACCACCCGCA TCACCGACGA GGGCTACTTC CCGGCCACCA TCACCGAGCA AGACGGCGAT
13681 GTCGACTTCC ACACGATCCT CAACGCGCCC AACGCGTTGG CGCTTGAGTT CGGCCACGCG
13741 CCGTCTGGCT TCTTCGCTGG CACCGACACG AAACCACCGG AGGCCACTTA CATCCTCACC
13801 CGAGCCGCCA TCGGCGGCAC CGTCTCATAA GGAGGTCACA TGGCGCGAAT GCCTCGCGTC
13861 CAGGCAGTAG CGGCCCCGAT CCTCCGGTCA GACCCCCGAC TGGAGGGAGT GACGGTCACG
13921 ACATGGGTTC CAGACGTGGA CTTCCGAGAG TTCCCGATGA TCAACCTCCG CCGCATAGGC
13981 GGGACGAGGA ACCCCAACGC ACCGACGCTG CACACGCTGC CGGTGGTCGA AATGACCGCC
```


SUBSTITUTE SHEET

```
14041  TACACCAGAG ACGGTCTCAT CGAGACTGAG GAGCTGTACG AGACCGCGCT AGAGGTTCTC
14101  TACGACGCGG TGGAGAACGG AACACAAACT CCCGCAGGGT ATTTGACCTC CATCTTCGAG
14161  ACGATGGGCG CCACTCAGTT CAGCTCCCTC TACCAGGACT CCTGGCGCAT CCAGGGTCTG
14221  ATCAGGCTCG GCGTCCGCAG ACCGAGAACC ACCCTCTAAC CGAAAGGTAA AGCCACATGG
14281  CTGAAAACGA CGACGCAGTG TTGACTGCGG CGGTCGGCTA CGTGTACGTC GGTGCTGCAG
14341  GCACCGCTGC TCCTACGCCG GCCTTGCTCA AGACCATCGA CCTCAGCAAG CCCGAGACCT
14401  GGACCGGTGC TACCGGTTGG ACGAGCGTCG GCCACACCAG CCGAGGCACG CTCCCTGAGT
14461  TCGGCTTCGA AGGCGGCGAG TCCGAGGTCA AGGGCTCCTG GCAGAAGAAG AAGCTCCGCG
14521  AGATCACCAC CGAGGATCCC ATCGACTACG TCACGGTCCT ACTGCACCAG TTCGATGAGC
14581  AGTCGCTGGG TCTGTACTAC GGCCCCAACG CCTCTGAGAC TCCTGGTGTG TTCGGTGTGA
14641  AGACCGGCCA GACCAACGAG AAGGCCGTGC TGGTCGTGAT CGAAGACGGC GACATGCGCC
14701  TGGGGCATCA CGCCCACAAG GCTGGAGTTC GCCGCGACGA CGCGATTGAG CTGCCCATCG
14761  ATGACCTGGC TGCGCTGCCC GTCCGGTTCA CCTACCTGGA CCACGAAGAC GAGCTGCCGT
14821  TCTCCTGGAT CAACGAAGAC CTCTTCAACG TGCCCGAGGT TCCCGAGGGC TGATCCCAAC
14881  TTGACAGCCA CCCGGCTGTC TACCCCGGAG GGGGAGGTTT CCTTGGCGGG CCTGGCCTCC
14941  CCCTCCTCCC GCCACTCACA GACCCGCCGA CACTGAAAGG TTCGCCATGA CAAACGTATT
15001  CACCATCGAC GCATTCCGCG AAGAGGTCAA GAAGAAGTAC GCTCCGGTCC TCATCGGCCT
15061  GTCCGACGAT GTGACCGTCG AGCTGAAGCC GCTGCTGAAG CTGGGCCAGA AGGCCCGCGA
15121  AGCGGTGGTC GAGGTGTTCA AGGAGTTCGC GGACATCCCC GACCTCGAAG AGGACGACGA
15181  CGACGAGTTG GTCGATGAGT ACTCGCTCCA GGTCTGCGAC ATCATCGCCA AGGCGTTCCG
15241  GCTGATCGCC ACGAAGCCCA AGAAGCTGAT CGCCGCCTTG GACGAGGAGC CGGATCCCCG
15301  TATCCGCGCA GAGCTGTATG CAGCGGTACT CAACACCTGG AAGCGAGAGA CGCAACTGGG
15361  GGAAGCCGCG CCCTCGCCGA GCTGATCGAC AAGTTCGGCG GGGCGATCCT CGCAGACCTG
15421  CTCCAGTACT ACCGGGTAGA CCTGCGCGAC CTGTTCCGCG ACGAGGATCC GCTTTCGCCG
15481  AGATTCGTTC TGTCCCTGGT GCTCTGCCTT CCCAAAGACG GCGCGTTCTA CGCAGAACGT
15541  CGTGGTGGGC AGCAGTACCG GGGCTGGACC GAGGACCGCT ACGCGCTCGC GGACATCTAC
15601  GACGCCATCC AGGCGGGCAA CCACATCCTG CTGCTGGCGA ATCGTGATCC GAAGAAGCCA
15661  AAGCCCAAGG CACCCAAGTC ATACCCGCGT CCCGACGACC TAGAGAAGAC CACACCGAAG
15721  CCGGGTTCGT TCGCCGCAAT GGTCGTGCGA GCGAAGAAGG CGGCTCGAGA GAGAAGGGAA
15781  AGGGAGGAGG AGAGTGCCGA ATAGTGCTGG CGTAGAAGTC GCCCGGATCT CGGTCAAGGT
15841  CAGCCCGAAC ACCAAGGAGT TCCGCCGGGA ACTCAAGACC GAACTCGAGA AGATCGAGCG
15901  GGAGCTTAAG GGCGATGTCG AGATCAACGG TCATCTCGAT GCGGCCCAGG CCAAGGCCGA
15961  CTTCAAGCGC ATGATGATGC AGCTCAAGAC CGAAGCTGCC AAGGGCGTTC ACGTCCCGGT
16021  CGACGTAACC GTCGACAAGA AGAGCAAGAA GGGAGGTCTC CTCGGAGGTC TCCTCGGCGG
16081  CAGCCGGGGG CTCGGAGATC TAGGCGATGA CGCCGAGAAG GCGTCGTCTC AAGTACAACA
16141  CCTTGGCAAG TCGTTCCTGG GCCTCACACG AGCCGCCTGG ATAGGCGTAG GCATCGTCGC
16201  CGTAGCAGCT CCGCTGGTCG GCATCGTGGC CGGTCTGCTG GCCGGTCTGC CGTCGCTGCT
16261  GTCTGCGTTC GGAGCCGGCG CTGGCGTAGT CGCGCTCGGC ATGGACGGCA TCAAGGCAGC
16321  CGCCTCGACG CTGGCCCCGA CGCTGGAGAC GGTCAAGGCC GCTGTCTCCT CGACGTTCCA
16381  GCAGGGACTC ACCCCGGTGT TCCAGCAGCT CGGCCCGATG CTGACCGCGA TCACCCCCAA
16441  CCTGCAGAAC GTGGCCTCGG GCCTCGTGAA CATGGCCGGG TCGATCACCG ACGTGATCAC
16501  CCAGGCTCCT GGTCTGCAGC AGATCCAGAA CATCCTCACC AAGACCGGAG AGTTCTTCAC
16561  GGGCCTCGGC CCTGTGCTCG CTACCGGCAC GCAGGCGTTC CTGACGCTGT CCAACGCCGG
16621  CGCGAACTCG TTCGGCACGC TCCTGGCTCC CCTGCAGGAG TTCACCAACG GCTTCAACGA
16681  CATGGTCAAC CGAGTCACGT CCAACGGCGT GTTCGAGGGT GCCATGCAAG GGCTTTCGCA
16741  GACGCTGGGC AGCGTCCTCA ACCTGTTCAA CCGGCTCATG GAGTCCGGTC TGCAGGCGAT
16801  GGGACAGCTC GGCGGTCCGC TGTCGACGTT CATCAACGGG TTCGGAGATC TCTTCGTCTC
16861  GCTGATGCCG GCGCTGACTT CGGTCTCTGG TCTGATCGGC AACGTCCTCG GGACGCTGGG
16921  CACACAGCTC GCTCCCATCG TCACGGCGCT CACGCCGGCC TTCCAGACGC TGGCGAGCAC
16981  GCTCGGCACG ATGCTCACCG GAGCCCTCCA AGCTCTGGGT CCGATCCTGA CTCAGGTCGC
17041  TACGTTGATC GGCACGACGC TGAACACGGC GCTGCAGGCT CTCCAGCCGA TGCTGCCGTC
17101  GCTCATGCAG AGCTTCCAGC AGATCTCCGA CGTACTGGTG ACCAGTCTGG CCCCGCACAT
17161  CCCGGCGCTG GCGACGGCCC TCGGCCAGGT CGCAGGCGCG GTGCTGCAGC TCGCTCCGAC
17221  GATCATCTCG ACGTTGGTTC CGGCGTTCGT TCAGTTGGTC CCAAAGGTCG CTGAGCTAGT
17281  TCCGACCATC GTCAACCTGG TCCAGTCGTT CGCCAACCTG ATGCCGGTGG TTCTGCCCCT
17341  GGCGCAGGCT CTGGTCAGCG TTGCTGGCGC GGTGATTCAG GTGGGTGTCT CCATCGGCGG
17401  CGCGCTCATC GGCGCGCTGG CGAACCTCAC GGAGATCATC TCCAACGTCA TCAAGAAGGT
17461  GTCCGAGTGG GTCAGCAGCT TCTCCAGCGG AGCCCAGCAG ATCGCTGCGA AGGCAGCGGA
17521  ACTGCCGGGG ATGATCCAGT CGGCTCTCGC CAACCTGATG GCCATCGGCC TGCAGGCCGG
17581  TAAGGATCTC GTCCAGGGCC TGATCAACGG CATCGGCGGG ATGGTCAGCG CAGCGGTCAA
```

```
17641  CAAGGCCAAG GAGCTGGCGT CCAGCGTGGC TGGTGCAGTG AAGGGCTTCC TGGGCATCGA
17701  GTCCCCGTCG AAGTTGTTCA CCGAGTACGG CCAGTTCACC GCCGAGGGAT TCGGCAACGG
17761  CATGGAGGCA GGGTTCAAGC CCGTCATCGA ACGGGCCAAG GATCTCGCGG CTGAGCTGTC
17821  CAGGGCGATG GAGTCGGGCA CCGACCCCTC CGGGATTCTC GCTGGGCTGG ATCAGAATGA
17881  GCTGAAGCAG ATGCTGGCGG CTCTCGAAGA GGAGCGCAAG CGACTCAAGG TCGAGAAGAA
17941  CGGTATCCCC AAGGGAGACA AGGCAGGCCG AGAGGCGCTG CAGAACCAGC TCGACCAGAT
18001  CCAGGCGCAG AAGGACATCC TGTCCTACCA GCGTGACCGC ATCAAGAACG AGTCTGAGTA
18061  CGGCGACATG GCCGGCGAAG ACCCGTTGGT GAAGGCAGCC TCCGGGCTGA TGAGCGCACC
18121  GGTCGACTTC GCGAAAGCGA CTGGCAAGCA GTTCCTTTCG GACATCGGCA TCAGCGGAGA
18181  TGGGTTCATC TCGAAGGCCA TCACCGAGGG CATCCAGTAC ATCTTCCAGA TCGGCTCTGT
18241  CGATGAGGCG CTGTCGATCA AGGACCGCGA GGAGTCGAAG AACGCGCTGT CCGTCGTTGG
18301  CCGCTGACTT GACATCCACC AGGAGGTAAG CATTGATCAC CGACACCATC GTTGAACTCG
18361  AGGGTGTCAA TGGTGAGCGT TTCAACTTGA CGACCGGTGA CCAGGGTGTG TACCTGGCCA
18421  CAGACGTGGA GGGTTGTTTC TACGACCCTC CCGTCAAGGT CGTTGTTGAA GAGCCGGGGA
18481  ACTACCCCGG CGCTCGCTAC TTGTCCCACC GAGCCCTGAA GCGAGACATC GTCTTTGGGG
18541  TCGTCATCCT CAACGACGCG AAGCAGGGGC CGCGCTCCTG GCTGTCGCGA GACTCCGAGT
18601  GGCGCAAGGC GTGGGCGTTC AACCGCACCT GCAAGCTCTA CGTCACCACC CCGGACTCCG
18661  GTACCCGCTA CCTGAAGCTG GCGCTGTTCG AGTCCCCCAC CGTCAAGATG GACACCGACC
18721  CAAGAGGTAA ACCCCTTGAG GTCACGGTGA TGTCGTGCAT CGCGTACGAC CCGTTCTGGT
18781  ACGAGGACGA CAAGGTCTTC TCGGCCAAGA CCAAGACCGA CACCCGGTTC GACCCGTCGT
18841  TCTGGACGCC GCCGTGGCCG TGGGAGGAAC TGCCCAAGGA GACGCTGCGG ATCAAGGTCG
18901  GCCGCGAGCA GGGTGGGCTA AACCCCACCG ACCAGTACAT CTTCCCGAAG TGGACCGTTC
18961  CCGGCTCCAC CGAGAAGGTG CCGAACTTCC CCTGGCCGTT CCCCCCGAAC GTCCCGATCC
19021  CGTGGGAGAC AGCACCGTTC ACTCAGTTCG TCATCCCGGA CTACTCGTTC GAGGATGAGG
19081  AGTTCCGCAA CCGCCGGCTC AAGACGCCGG GGTTGATCTA CGGCGAGAAC TGCGTCATCG
19141  ACACCGACCG GCGCGAGGAG CAGATCGCTT CCGAGTCGGG CTCCCCGGTG TGGGCTCGGA
19201  TGAACGGTGT CCGGTTCCGC AACTCGATCC CGCCCTACAC CGAAGAGGCT GAGTTCGTCA
19261  TAGACGCATC GGGATGCGCT CCGGGACAGG TAGTTACCCT CCGGCTCACG AGGCCGTGGT
19321  CGCGCTGCTG GGGGCTAGAG TGAGTGGTCT GACGAGCGTT CGTGAGGCCG AAGATCTCTG
19381  GCAGAAGATC CAATTGCGGC GCTGCAAGCG CGAGCAGGAA CGGCTCAAGC ATCCCGACGT
19441  AGAGCTGCGC GATGGCGACT TCCGCCTGCG CGGCCTGGTC GCTGGCGAGC GGGTGCTCGA
19501  GTGGGAGTTC ATCGAGAACG AGACTGGCAC CTGCACCTTG CAGCTCTCAC TGAGCCATTA
19561  CCTGGCGAAG TGGGTGATGG ACCACCGGGG TCGAGCAAAG CGCAACGTCA TCATCAACAT
19621  CGAGAAGCAA GGCGCTCGAT GGACCGGGAT GATGGACCAC TACCGGGTCA TCAAGACCGA
19681  CGCAGGGGAC GCCTACATCG AGATCGTGTT TTTGCACGAC TTCGAGCAGA CCAAGCATAT
19741  CCGGGTATGG TGCAACCCGT TCCTACGCCC CGAGCTGCAG TTCCCCAAGG TGTGGATCAT
19801  CTTCGGGCCG GCCAAGTGGT GTTTGCTGGT GACACTGTTC GTCAACCTGC TCAGGCTCGA
19861  GACGAGCTTG TGGACGCTGC CTGATGACCC CACGGACATC AACGAGTGGA TGGGTCCGAG
19921  CTTCAACCCA GCAAATTGGC GGAACATCGT CAAGCCGTTC CCGTTCCTGG CCGACAACTC
19981  ACCGGTCACG ATGGTGTTCA GCCGGTTCGG GACGTTCTAC GACACCGCCA AGAAGATCCT
20041  CGAAGACCAT CAGCTCACGC TGACGTGTCG TCGGTACATC AAGGACCGCG ACCCGCATCC
20101  GTTCGAAGAT CTCAAGGGGC TCTGGGGAAT TGATCCTGTC GAAGACCTGC TGCAGAAGAT
20161  CCCGCTCCGG GACGGCTGCG TGGTCTGGGA CATCGAGGAC AACTCAGGTT GGGGCACTCA
20221  GACCGCGTTC GGCGGTTCGT GGCTGACCGG GTTCGTCCGA GGGATGGTCC AACTGGCCGG
20281  CGACGGCCAG GTCGAGGGCG TCGATGTGTT CACCGGGGAC TACACGTTCC CAGGCGAGTA
20341  CTACTCCCCC TGGTTCATGG GCACCAGCCC GATAGCACCC CACGTCGTGT TCGAAGAAGG
20401  ACCGCTGACC GGGATCAAGT CGTCGGAGTT CTCGTACTAC GAGGCCACCG ACACCAGCTT
20461  CCTGGCTGGT GGACAGAGCG CACCTGGCAT CAACGAGGGC ATCTCGGCCC TGGTGAACAT
20521  CGGTGGCGAC CTGCTGACCT CGTTCATCAA CAGCCAGCTC GCCGCGCTCG GCGCGGTCGG
20581  TGGAGCGATT GACCTCCCGC CTCTGGGCGG TCTGCTCGAT GCGGTGTTGC AGCCTCTGTA
20641  CTCCGATGTG TTCGGCGCGT TCATGGAAGT TCCGACTCTG CGTGCGATGG GCATCTCGCT
20701  CCCGATCTCC GGGCTCGAGG ACATCGTCAC CGGACTGGGC GACTTCCACT ACTTCGAGAA
20761  CATGGCCGAC GGGGCGATGA AGGCGTTCAC GCTGTCAGCG TTCGCAGCCA TCGCATCGCA
20821  GATCCACAAG ACGAGGGCTC GAACGACCCA CACCCTCAAG GTGTCTGACG CCGCTCCGTA
20881  CATCTTCGCG CCAAAGCCCT ACGGGCACTG CTGGATCGGA GATCGCGTCG GCACGTCGGT
20941  CCTCGGCTAC CCGGTCGAGC ACCAGTTGTT CGTGGAGCGC ATCCGCAAGG TGAAGTACCG
21001  CATCGACAAA GACGGCATGA AGCCGTTGGA GATCGAGATC GGTTACCGCG AACCGAAGAA
21061  CCCAGCACTA CACATCCTCG AAGAGATCAA GCGCGTCAAC GGCGCTCTTG GCACTGCGGG
21121  GATTCTCTAA ACCGAAAGGC ACGCCGCATG ATTCCCTCAC AAGAGTCTCA CAATCCGAAC
21181  GACCCGCGAC AGCACGTCAT GTGGGCGCTA CGCAATCTCC CGATGATTGC TGGCGTCGGG
```

```
21241 GCGATCACGC ATCCGGGTTA CCTGGCGGAT TGGTCAGAGC ACTTGTGGAA GTGCGGCTTT
21301 CGGCACGTCG ACTGGCTCCG GGAGCTGGCT GATGAGGACG GCAACATCCA CGTCAGTCAG
21361 CTTCCTGACC AGGAGATCAA GTTTCAGCAG CCCTTCCGGG GCCAGCGAAG CGACTACAAC
21421 AACGCAGCTC GATGGGTCGG CAAAGACGAT CCTGACCCAG AGCCCGTGCG TATTCCAGAC
21481 ATTCGCAAGC TCACAGACCA GGAGAACAGA GCGATGATCG CGCAGTACGA ACGAGACGGT
21541 TGGATCAAGG ATGGATCCCC CGGCCCAGCG ATAGCCGAGG TCGTGGAGTG ACCCCGTTCA
21601 ACCCAGACTC CATAGGCGAC TACGTGACAC TGCTCGGCGT TGCGTTCCTG ACCTTCTCGG
21661 TTCCCGCATG GTTCACCGGA CGAGCACGCA AGCACAGCAG TGACATCGGC GAAATCAAGG
21721 AACAGGTATG TAACACCCAC GACACGAACC TGCGCGATGA CCTCGACAGC GTCAAGGCAG
21781 ACATCAGCGA CTTGAAAGAG ATTGTGTTGC AAGGGTTCCA CCAGGTGAAC GAGTCGATCA
21841 ACCTCGAGCG CCGTGAGCGG ATCGAAGGAG ACCGCCGAAA GGAGGTTGCG TGACCTACCC
21901 CACCAACCCA CTAGAGGCCA TCGGCGCTGA CGGCGCATTC GAGATCGGTG GGGGCGACTG
21961 GAGCTTCGGC CAGGACTACA CCGAACAGGC CATCCGGGCT CTGTTCACGA TGCCAGCGGT
22021 CACGATGGAG AACGCTCTCG GCCTGCTCGA AGAGCACCTG CTGAAGCTGC CTCTGGAGGC
22081 GCTGCAGGGC TTCAAAGACA TGATCCCGGA CTGGGTCGAA GGAGCATTCG ACACGGTCAC
22141 CGGCGCTGTG CAGGCGATCA TGAACGCGCT CCAAGACGGC CCGCTGTTCC TGAAGTTCGC
22201 CGAGTTCCAG CTCTTCCTGC AGCGTCTGCT GAACAACCCG GCCGAGGTCA TCGGCGAGAT
22261 CCCCCAGACG TTGATCGACG GCCTACAGGA CGCGCTCAAC ACCGTCAACA ACACCATCCA
22321 GACCATCGTG GACATGCTCC TGCAGGCGCT GGGCATCACC CCGGAGGGGG AGCTGATCGA
22381 CCGGATCTTC GACCTGAGCG ATGAGATGGA GTGGCTGCAG ACCGCAGCCT CGAATGCAGC
22441 TACCGGCATC CAGGACACCT GGAACAAGTT CTGGGGAGCC CTCACCGGGC GCGTCCCAGA
22501 CCAGGACCAG ACCGTCGCTG AGCCCGCCGA GCGTATCGGC GAGCTGGCCG GCACCACGTC
22561 TGCTAACTCG TCTGCCATCG CGGAGCTGCA GCGTCGACTG GACAACCAGC AGAACGCTGG
22621 CGGCGTGGCC GGCGGTGACG ACTTCGAGCG ACTGAACATA TCCGGTTGGG ACATCAGGTA
22681 TTCCAACGGA TCCAGCGGCC GAGGGTACTA CCGTGCCGAC GGCCACCAAC TGGTCTGGAT
22741 GGACGAAGGC AACCAGCAGA ACACCGCGAC GTTCGTCCGC ACCAACCCCG CAGACGAGAA
22801 GACAGCCACC GACTACCAGA AGATGACGTT GGTCGTCGGG ACTATCTCCG GTGAGGTACA
22861 GACCGTGTTC CCGCCGCAGG GAGGTTCGCA CACCCGGCTA TGGGTCCGCG TCAACGACAA
22921 CGCTCCGACC GTCGGCATCA CCGACGGCGT GTTCGTAGAG ATCGGCGGCG TATCGAAGGC
22981 CCAGATCGGC TACCGCCGCA ACGGCAATGA CACGTTCGTC GGATCTATGG TCGACTGCAC
23041 CTGGGGTGCT GGATCGATCT TCGCTCTGAC CGCCGGCACG GCCAACGGTG CTGAGAAGTT
23101 CGAGGTCTCG AAGAACGGCC CCGTGCTGGC CACATGGTCG GACGACGGCG TCGTCTCCGC
23161 GATGGGTGCG AACTACCGCC GCTGGGGCTG GGAAGGCCAG GCTCGTAACC GCAACCTCGG
23221 CCAGGGCACT CCGAACTCGG TCACCCGAGT GACGATCACC GACAACGATC CTACCGGCGC
23281 AGGCGGTGGA GCTGTCAACG TCGGAGGAGA TGTCGTAGGT GTACTCCCCA TAGAGAACGG
23341 AGGCACCGGA GCTTCGACAG CTTCGGCAGC CCGTACCGCT CTCGGAATCG ATGACCTGGT
23401 CGAAGATATG TCCGACGTAG TTCGTGGATC CGTCGAAGGA CTCCCGTTGA TACCGAAGAT
23461 CTGGGTAGGA ACAGAAGCTC AGTACACGGC TCTCGCCACC AAGGATCAGT CCACGCTATA
23521 CTTCAGGACC GCTTAATGAC TGGTATCTCG TTGGGTGTCA ACGACATCCG CAACCTCTCG
23581 ATATTCTTAG GCGTCAGCAA CAAGATATTG AAGGTCAGTC TAGGCACAGA AAAGGTCTGG
23641 CCTGCGTTCA CCCCGGTGCT GACCACGTTC GCCACGGTCG GCACGTACAC CTACAACATC
23701 CCCGACGGGG CCAAGTTCAT CGACGTCATC CTCCTCGGAG GAGGCGGCGG GGGTAAAGGC
23761 ATGGCCCTGG CTGACGGCTG GGGCAGAGGT GGAGACGCCG GAAGCTGGGC TATCGTCACT
23821 CTCGAACGCG GGGTACACAT CCCGTTGTCG ACCAAGACGA TCACCGGGCT CGTCGGAGCT
23881 GGAGGCGCAG CGGGAGCTGG CTCTGTATTC TCAGGCAAGG CCGGAGGCCC TGGAGGAAAC
23941 ACCACGGCGT CCGCTGTCGG ATGGTCAGGT TTGACCGCAA CCGGCGGTCC CGGAGGCTCT
24001 GTGATCGACA TCCTCAGCGT CGCCGGAAAG TCGCCTGGAG ATCGGACCTA CAACGACCAG
24061 CTCTACATAG GCGGCGCACA ACAGAACTCA GCTGGCGGGA ACGGCAATGC TCCTGGCGGC
24121 GGCGGGGCTG GTGCCCAGGT CTCCGCACAG AGCGGCGGTG CTGGCGCTCG CGGCCAGGCG
24181 TGGTTCTTCG CGTACTGACA AGAAACCCCC CTCTTTAGGA CTCAGTGTCC TTGGGAGGGG
24241 GGCTTTTTGC GTTTCAGGAG GTCTTGGCCA GCTTGGACAT CGCCTCAGCG ATAGCCTCGT
24301 CGCGGGCCTC AGACGCCATC TGGTACTTCA TCGCCATCCT AGGAGTCGTG TGACCGAGAC
24361 GGGCCATCAG CTCCTTGGTC GTCGCACCTG CCTGAGCGGC GAACGTAGCG CCGACAGCGC
24421 GGAGGTCGTG GATGCGGAGT TCCGGCCGAC CGATCTTGGC GTAGCCACGC TTCAGCGACT
24481 TGGTGAACGC GGACTTCGAC AGCCGGTTGC CCTGCGTCGT GGTCACCAGG AATGCCTCGG
24541 GGCCCTTGTT CATCTTCGTA CGGTCCTTCA TGTGCGCTCG GATCATCTCC GCGACGTGAG
24601 GCGGAACCGT CACAGGACGC TTCGACCGGA CGGTCTTGGC GTTGCCAACG ACGATCTTGT
24661 TCCCCACGCG GGAAGCGCCA CGGCGCACCC GGAGCTTCAT CGTCATGCCG TCGTCCACGA
24721 TGTCCTTGCG GCGAAGCTCG ATCAGCTCTC CGAACCGGAG GCTCGTCCAC GCCAGGATGT
24781 ATGCCGCGAT CCGGTAGTGC TCGAAGATCT CAGCGGCGAC GATGTCCAGC TCCTCAGGCG
```

```
24841 TCAGCGCCTC TACGTCGCGC TCATCGGCTG CCTTCTGCTC GATCCGGCAC GGGTTCTCTG
24901 CGATCAGCTT GTCCTCGACC GCTGTGTTCA TCACCGCCCG GAGGACGTTG TAGGCATGCC
24961 GGCGGGCAGT CGGGTGCTTC CTACCCATCC CGGCCCACCA CGCACGCACC AGAGCTGGCG
25021 TCATCTCTGT GACCGCCACT TCACCTAGCA CCGGGTAGAT GCGGCGCTCC GCGTGCCCGC
25081 TGTACAGATC CCTGGTGCCG TCTGCGAGGT CGCGCTCCAC GAGCCACTTC CGGGTGTACT
25141 CCTCCAGCGT GATGGCGCTG GCGGCTGCCT TCTTCGCCCG GTCCTGTGGA GGGGTCCAGG
25201 TCTCCATCTC GATGAGCCGC TTCTCGCCCG CGAGCCAGGC TTCGGCGTCC ATCTTGTTGT
25261 CGTAGGTCTG CAGCGCGTAG TACCTCACAC CGTCCTGCGG GTTGACGTAT GAGGCTTGGA
25321 TCCTCCCGCT GCGCTGAGTC TTCAGCGATC CCCATCCGCG ACGTGCCAAC TAGGTCTCCT
25381 CTCGTCGTGA ACAAGGCTAC CGGGTTGCAA CTCCTGTGCA ACTCTCAGGC TTCAACGCGC
25441 TTCTACGACC TGCAATTTCT TTCCACTTAG AGGATGCAGC CGAGAGGGGG TAAAAACCTA
25501 TCTTGACCGG CGCATATGTG GTCGGCAGAC ACCCATTCTT CCAAACTAGC TACGCGGGTT
25561 CGATTCCCGT CGCGCGCTCC GCTGGTCAGA GGGTGTTTTC GCCCTCTGGC CATTTTTCTT
25621 TCCAGGGGTC TGCAACTCTT GTGCGACTCT TCTGACCTGG GCATACGCGG TTGCAACGCA
25681 TCCCTGATCT GGCTACTTTC GATGCTGACA AACGAATAGA GCCCCCCGCC TGCGCGAACA
25741 GACGAGGGGC ATTCACACCA GATTGGAGCT GGTGCAGTGA AGAGAATAGA CCGGGACAAG
25801 GTTGCACCGG GAGTTGCAGC GGTCGGAACC CTCGCCGTCG GCGGGCTGGC GTTCGCCCTG
25861 TCGTTCACGG CTCTCAGCGA GCTGGCTGCG GCCAACGGGG TGGCCCAAGC AGAGATGGTG
25921 CCCTTGGTGG TCGACGGCCT GACGCTCGTC GCCACGGTCG CCACAGTGGC CCTCAAGCAG
25981 AACAGTTGGT ACGCGTGGTC GCTGCTGATC CTGTCCACCG TCGTATCGGT GGCCGGCAAC
26041 GTGGCACACG CCTACCCCCA CGGCATCATC GCGATGGTGA TCGCTGCGAT CCCTCCGCTC
26101 TGGCTACTGG CGTCGACCCA CCTAACCGTG ATGCTGGCGA AGCAGCACTC GGAGCACGCC
26161 GAAGTACCTG TCTCGCGGCC AGAACCCGCG CCTCGGGGCC TGGAGCCCGC TGCCGCTTGA
26221 CTGCGCCCGA CCGGGACAGA AATACATAGA GAACCTATGG ATGTAGGAGG CACAAAAAAA
26281 TACCCCCCGA GCCAGCCCGA AGGCCAGCCC AGGGGGCATG GTTCTGCTTC AGTAGACCTT
26341 GCGAGTCCGA CCCGAGTTGA TCATCGCCAT GATGACCCAG ACGGGCAACC ACATTCCGCA
26401 GGTGATGAGC GAAAGCAACA GGTGCATCGC GTGGTTCGTC CTGACAGGCA TGACAGTGGG
26461 CTGCGGCATC GGAGGAGGCG CGACCGGGTA CGGCGAGCCC GCGTACCACT GAGGTCGATC
26521 TTGTTGGGGC GGATACTGAT TGGTCATCCC GACAGCCTAC TTGCCGATGG GTCGCATCAG
26581 CTCCTCGACC GACTCGCGCT CCACGCGGAT CAGCCGGGGA CCGAGCCGAA CGGCCTTGAG
26641 CCGGCCGTCG GCGATGTAGT TGCGGACGGT CTTGGTGCTG ACACCGAGGT AGTCAGCGGT
26701 CTCCTGGATG GATGCTCTCG GGGGCATCAG CGCGGTCCTC CGTGCTTCAT CGGTTGTCTC
26761 CCGAACCCTG GATCACGCCA CGATCCTTGC GGCTCTGGAG CTTGTTGAGG TTCCTCTGGG
26821 TGACGGTGCT CAACCAGACA TCGAGCTGGT TGGCTAGCTG GGCGACGTAC CACATCACGT
26881 CTCCGAGTTC CGCCTGGAGG TCGTCTCGGT TCTCCTGGGT GATGACACCG TCTTTATCCC
26941 GGAGGATTTT CTTGACCTTG TTGGCGATCT CGCCGGCTTC GCCTACGAGA CCCATCGTCA
27001 CGTAGGAGAG ACCCTCGATG CTGTCGCAGT CGCCTGCACC GGGGTAGATC GCTGTGTCGC
27061 TCGCGGCGAT CTGGTAGATG TCGACGTGCA TCAGATCATC ACCGGGAACA ACTGGCCACC
27121 GGGCATCTGG ATGAACACCG GGACGCTGGG GGTGTAGTCC GACGAACCCG TGCCGCCCTC
27181 ACAGGCGGAC AGGCTCAGGG TGGCGGCAAG GCCGATGATG GCTGCTGCGA TGGTCTTCTT
27241 CATCTGTTGC TCCAGTAGCT AAGTTCGGAC TCCAGTTCGC GGATACGCTC CTGTAGCCCT
27301 TGGTTTTCCA GGTACGCCTC GGCGAGGTTG GCCTCGGCGC GGTCACGGGC CTCGTCCTTC
27361 GACGTGGCCT CATCGATTGC CTCGTGTAGC CGGCGGATCA GATCTGGGAT GGCACCGTGC
27421 AGACCGCATA TGAAGTCGGC GTCTGCCTCG GAGAGGTGGG ACGCCACCAG ATCCTTGTCC
27481 TGGGTCTCCT GGTTGACCGC CCAGATGACG TGATCCTCTA GCCCGTGGTC GGTCTCGCAG
27541 ATAGAAGGCG GTTCTACCTC CTCTGGCATC CAGTAAGTCT TCTCAGCCCC GGTGGACTTC
27601 GCCCACTGCT GGTAGAGGAT GTCGAAGAAC TCGTGGTCCT GTTCGTCGGC GGTAATCACA
27661 GATCGTCCTC TTCATCCCAT TCGTCGTAGT AACACGTACA GCCGCAGCAG GTGCAGCAGC
27721 CGCACTCGTA GGTGCCGTAG TCGTAGTCAT CCCAGTCGTC TTCGTCCATC TAGCTGTACT
27781 CCTTCATGAT TCGGTCGAAC GCACGCGTCT GCACGCGCAT CTCCAGGTCG ACCGTTCGCT
27841 TCAACCACGC CCATTCGCCG TCGTGGTTGA TCTCCCACTG GCTCTTGAAT GTCGCTGTCT
27901 CAACGAGGAA CTCGACAGTC AACGTGTGCA GTCCGTTGTT GCTGGGCTGG AATCCGATAC
27961 CGTCCTCAGC GATGTACCAG GGCAACTCCT GGCCGTCGAA GTAGACGGCC TTGTCGGTCA
28021 CCAGTACTTC AGGGAAGGTG TGCTCGGTCA ACGGCGTCCC AGGTATGGGA TGACGCTGGC
28081 CCGGAACTCA AGGAACACCA TGTTGTCCGG GCAGTCCTCG GGGACGTTGT CGGGGCGTTC
28141 GGCGGTGTAG ACGCCGATCT CGTTGCCCTC CAGGGTTCCA AGCTCGTTGA GCTTGTAGAT
28201 CGCCAGACCC ATCAGCTCTT CATCGAGACC GTTCGGTGCT GGCAGTACAA CTTTGGCTTG
28261 TGGCATTAGC CCTCCCTCGG AATTACGTAT GCGCTGAACT CGACGGCCGT AATGCCGTCT
28321 GGCAGTTGGA ATCCGAACCG CTCTTCGAAC TCCTCGTTGG TGATGGGGCC GTACTCGAAG
28381 GTTCCGGGCA CTACCTCGCC CTCCCCCTCG ATCAGGAGGT ACGCACCGGC GGCGTACACC
```

```
28441  TCCTCGTCGT TCGGCCATCC GACTACGGTC CCGAGGACCG TGAACTTCCT CGGCTCCATC
28501  AGGGCACGTC CACTTCGTTG ATGAGGAACC GCATCGGAGG TGGAGTGAGC ATTGCCTCGG
28561  CTATGGCGAT GAGGGCGTTC AACTGACCCT TCAGCAGCTT CTCCTCGTCG CCTGCGGGAA
28621  GGTGGCGCAC TCGGCGCTCC ATCTCCTTGG CGCGTTCCAG ATATTCGGTG GCTGTCAAGT
28681  TGTCCTCCTT AGTAATCAGC GCCGTAGAGC GAACCCCACG AACGCTTTCC GACCTCGGGG
28741  TCGGTGCCAA CCAGCACCGG ACCCATCTGT TCTTGCATCA GGTGGCCAAT GTGTGCAGCG
28801  GCTCTCTCAG CCTCTGAGGC GGGCAGAGAC GCGACGATCT CGTCGTGGAT AGGCAACCGT
28861  AGGTACGGGG TGTATCCGGC CTCGTGGAGG CGAATCAGAG CCCGACAGGT CACGTCCCGC
28921  GACGACGACT GGATCATGTA GTTCAGCGCG GAGTATGTCC GCGAGCTGTC CACCGGCAGC
28981  CGCCGGCCCA TCGCGTTGAC GATGTAGCCG TTGCGGCCAG CTTCCATCGC CAGCTTCTTG
29041  CTCAGCCGCT CCACACCGGG GTATGTCGCA GAGAACGCCT CATGAACTCG CTTGGCCACA
29101  GGGATCGAGA TCCCCACTGC CTCAGCGAGA GCCTTCGCCC CACCGCCGTA GACCTTCTGA
29161  AAGTTGGCGG TCTTCCCAAC CTTTCGCGGC ACCTGGGCTG CGTCAGCGGT CATCTGGTGG
29221  AGGTCCGCAC CGTTCTCGAA TGCCTCGATC ATGTTGCGGT CGCCCGACAG CGCCGCCAGG
29281  ACGCGAAGCT CCTGCGCCTG GTAGTCGACT GAGGCCATCA CATCGCCTGG CTCAGCGATG
29341  AAGCATCGCC GCACGATCCA GTCCGACGAC GGCAGCGTCT GCGCCGGGAT GCCGGTGATC
29401  GACATGCGCG AGGTCCGCGC CTGCAGTGGG TTGATGAACG TGTGGCAGCG GTCCTCAGAG
29461  TCCCTGGTGT CGATGAACTT CTGGACCCAG GTCTTCCGCC ACTTCCCCAG CTTCTTAGCC
29521  TCCTGAGCGA TGGCGGCAAG CTCGTTGCCA TCTTCGACCA GCTTGTCGAG CAGAGCCGCG
29581  TTGACCTGGC GCTTGCCAGT CTCGGTGCGA CCGGTGATCT TGACGCCCAT CTCCTCAAGC
29641  CCCTCGGCCA GATCCTCGGT CGAGTTGACC TTCTCCACGC CGTACTCGGT GAAAGCGATT
29701  GCCTCCCAGA CCTCCTGATC GGCCAACCAC TTCTCGGCGA GCGACCGCGA GTACTCCACA
29761  TCGAGCAGGA AGCCCTGCCT GTCGATGTAG CTGCAGATCT CACTGATCTT GTGCTCGTAC
29821  GGCACCAGCG ACCGACTCAC GTCGGGCACC AACGGTGTCA GGCTCTTGCA GACCCTCGCG
29881  GTGAAGATCG TGTCCATCCC GGCGTACAGC AGGTACTCCG GGTGGAACAG GTCGATGGTC
29941  GACCAGATCT TGGCCTTGGT CGTCTTGTGC TCGGCGGCTA GCTTGGCCAT GAGCTTCTTG
30001  ACGTTCTCGG CCTGGTCCTC GGAGATGAAC TTCGCGATCA GCTCTTCGAG CGAGTGCCCG
30061  AACCCGCCGG CCTCGAAGGG CCGGGGGTCC ACCAGCTTCG CCAGGATCTG CGTGTCAAGC
30121  ACGCGGGGCC ACAGACCCTC CATCTCGATC CCGAAGCACT GGTCGAGCAC CTGGAGGTCG
30181  AAGGAGGCGT TCTGGAGCAC CATGCGCTTG AGAGCGCCGA TGGCGATCCG CACGTCCTCG
30241  ATGAACACGT CTCCCAGCTC CACCGGCACC ACCCAGGCTT CGTCCTGAGT ACCGAACTGG
30301  ACGAGGCGGC ACTCGAAGGT GTCGCTGTAG ATGTCCAGCC CGGTGGTCTC AGTGTCGACG
30361  GCGAGGCAGT TCAGGTGAGC CCGGATGAAG TTGCGGAAGC CTTCCAGATC CTCTGGGGTT
30421  TCAACGACGT TGACGGTGAC GAGGTCTCCC TGAACCTCAT GCCGCAGCTC GATCAAAATG
30481  CTCTCCTACT GGAAGTACTG AGGCGGAATC CAGGTGGCTG AGGCCATCTC CTTGATGGCC
30541  TGCTGCATGG CCGCTTCGAA CGGACAGTCC GGGTCGATGT CCGGCTTGTA ATGGGTGACG
30601  ATGATCCGGC TGTTGCCGCC GAAGTCGTGG CTGACCAAGC CCTTTGGGGG CAGCTTCTTC
30661  AGCGCCTTGA TCAGTTCCTC AACCGTGGTC CCGGTAGGGG CCTTGCCGTC AGGCAATGCC
30721  TCCCCTCCGT ACGGCACGTC CAATGGGATC GTGTACCGCT CAACGTCTTT GATCTTCATC
30781  GAGCCTCTTC CTCTTCGACT ACCTCGTCTA CCCGGCGGAA TAACTCCGCT AGTTCTGCGG
30841  GTAGCAATAC TGGGTACTTC TCTCGGGCTT CCTGCATCGC TACCGCGATC CCAATCAGGG
30901  CAGCGAGCAG TTCATTGACG GAGTACGCCA ACAGCTCTTC GCGGATCTCT TCTCGGGTCA
30961  TTAGTGGTAG ATCCCCCGGA CGGTGCGCGA GATCGTGGCA GGGTTCACGC CGTAGTTCTC
31021  GGCGAGATCC TTCTGCTTCA TACCGCCCAG GTACGCCTGG CGGATGTCCT TGACCTCGCG
31081  CTCGGTGAGC TTCTTGCGGT TCGGCCGGCT CGGGCCGGTC TCAGGCTTGA CCTGAGCCAG
31141  CGCCTTGCCG AACAGCTCGT TCTGCGTCCG CTGCTTGATC GCGTACCGAC GGTTCGCTGC
31201  AAGCACCTCG TTGAGCCGCT GGGACAACTT GACATTGGCC TCACGCACTA CCTCGACCTC
31261  TCCGAGCAAG TTCGTGATCC GGTAGTCCTT GTCCTGGTTC TCGATGGCCA ACCGGTTGTT
31321  CTCCTCGGAA AGCATCGAGA CCTTGTATTG CGCCTCTCCC AGCGCAGCTT TCAGGTGCTT
31381  CTTCCTCATT CAGCGCCCCT CTCTCGGCGG AACTGTTCGT ACTCGTCTTC GGTCATGTAG
31441  TAGTAGTAGT CAACGACCTT GTCCCAGTTG AAGGTTCGGG ACGTGCCGTC ATCGAACGCG
31501  ATGATCAGGA CACCCTCTTG GGTGTCTAGG ATCGGCTCGC CAGCCACGAC GTGGAAGCGG
31561  TCCTCGAGGG TCACCGCAGT CGCTCTGCGT GCCATGTCAG TTCCTCTCAG TAGCTGTAGG
31621  GGACATCCGG GATGTCCTGG TAGGTGTTGG GTGCGATCTG TCGGAGCTGC CGAAGCAATT
31681  CCCCTGCCAG CTCACGGATC TCGGCATCCG CGGCCTCGTG CCAGCGGGCC TTGATGACGT
31741  ACCGCCACGC CCGATGGTTG CCCGTGACGA CCATCGGTGA GTTCGTCATG TTCGGCAGGA
31801  CAGCTCGCGC TGCCTCGCGG GCCTGCTTGC GCGGCAAGCC CCGGTCAGCC AGCCGGTTGA
31861  CGATGTGTTC GTAGACAGCG TCAATCTCAG AGCTGACGGA CTCCATGATG TGGACGAGGT
31921  CGTCTCGGTC GTCGGGGTGG AGCTTGAACA GAGCCGGGGG CAGATGGATG CCAAGGTCGG
31981  TCGGATCCAC ATATCGCTGA GACACCACCG AGAAGCTCAA GTGACGGTGA CGCTCCAGCT
```

```
32041  CGGTCAGCAC CGACCTGCTG GCCTCGATGT AGAACGTCGC CGAGGCGTGC TCGAACACGC
32101  TCTCGTGGCC CAGATCGATG ATGTGGTTGA GGTAGTCCTC GTTCTCGGCA GTTGCCGGGT
32161  TCGGTCGGTG GAACGACCGG TAGCAGTTCC GGCCCGCGAA CTCGGCCAGC TCGTCGGCAT
32221  CGAAGTCGCC GAAGTAGGGA TCTTCGTCCT TGGATTCTTC GAAGTCATCG ACCTCGAATC
32281  CGATGTCCCG CAACGCACCC GGATCGATCT CGGTGGCAGC GATCAGTTTG GCTTTCATAC
32341  TCTCCGCTCA GAGTTGGTGG AACGAGGTCA GCCAGGGGGC AGCGAAGCCC TTCTACAGCT
32401  CCCCTTGGCT CGTTACCGGC TTCTCGACCT CGGTGGATGT CAAGTAGTCG AGATGACTAC
32461  TTCTTGTCGG GCCATTGCGC GTCACACTGC TGATCGCGAG GTGCGGTGCA GGAGAACAGC
32521  GCGTACGGCT TGCCCGTCTT CTTCGAGACG CCCGACTTGT AGACCATCTC GCCGTGCTGG
32581  CAGTACCGCT TCTCGCCACC AGGCGCTTCC TGAGCTGCCT GCGGGGCGCG AGACTGCTGC
32641  TGGCCACCGC CGCCGCCGTT GGCCGGCGCG GATCCACCGG AGCCTGCGTA GTGGCCTGCG
32701  ATCTGCTGGA CCTTGTCCAT CAGCGCCTTG AACTCGGCGG TGTTGACCTT GGCCAGCACG
32761  TCGGCCGGGT CCGCACCCTT CACGACCACC CACGGGTCGC TGTACTGACC GGCGAACTTG
32821  AACGTGGCCG ACACCCCATC GGTGGAGTGC TGGACCGCCA TCGAGTCGCG CACAGCAGCC
32881  GAGGCCGTCG TCACCGTCGC CGACGGCGCG GTCTCAGGCT CAGGAGCCGG GGCCGGCTCG
32941  GGCTGGGCAG GGGCGGTGCT CCACGGATCG TCGTAGGACA ACTGGTTACC TTTCACTTAA
33001  TGGGGCATGC GCCGTTGGCG CACTCTTCAT CGACACCGTC TTCGACGGCT TTGGCCGCAG
33061  CAGATTCGTA CTGCTGCTTG GTGATTCGCT CGTACGGAGC CTGCGGGAAG CTGGACTCCG
33121  GGAAGATCGT GGAGCCCTTG ATGAGCCCCG CGAACCTCTT GAGATCGGCT GCGACATCCT
33181  CGGCCTCGTA GGCGTCTGGA TGGACGTTGG CGGTGAACGA CACCGCGTTG TCAGCCCAGC
33241  ACATCTGGTA GAGCGCCTGG AACGCCAGGA GCTGGTGGAG GGTCAACTCG TCGGCTGACT
33301  CAACGATCTC CTCGTCCCAA CCGAGTTCCT CGACAGCCTG GACCAACGTG TCCTTGGTCG
33361  GGATCGAAAC CACCTCGGTG TTCGGAGCGA AGAGATCCTT CTCGATCTCG TAACCCTCGG
33421  CTGCCAACCT CCGCAGCTCG GCCATGTCGC TGTTGAGGTT GAACCGCACA CGCCGGATGA
33481  AGTACCGCGA GAAGATCGGG TGGATCCCCT CGGAGACTCC TGGCATCTTC GCCACCGTGC
33541  CTGTGGGAGC GATGGTTCGC TTCTTCACCG GGACAGGGAT CCTCAGATCA TGGGCGAACC
33601  GTTCGGCCTC TGAGTCGACC TCAGCGGCCA TCTCCCGCAA GAACTGGGTG AACCGCTTAT
33661  CTCCGGGTGC CTCGGAGTAC CTGCTACCTG TGAGGGCCAA ATAGGAGGCA ACTCCGAGAT
33721  GACCCACGCC GATGCGACGG TTTCGGTCCA GAACCTCCCG GCTCTTCGGG TCGGCCACTT
33781  CCGAGAACGT CGCCCGGATC AGGAATCTCG TCATCAGACG ATGCGCCCGG ATCAGGTCGA
33841  GGTAGTCGGT CTTGCCGGCC GGCGTCACGA ACGCCGCCAG GTTGATGTGG CCGAGGTTGC
33901  ACGGCTCCCA CGGTTCGAGA GTGATCTCGC CGCATGGGTT GGTGCAGACC ACCCGGTTGG
33961  GCTCACCGAC GTTGGACAGT GACGAGTCCC ACATCCCCGG CTCTCCGTTG CGTACGGCTC
34021  CCTCGGAGAG TGCCTTGAGC ACTCGGTGGG CTCGCTTCTG CTTGGGCATG TCCTCGCGGG
34081  CGACCGCGAA GCTGCCGTAG CCCTCCTTGG CCAGACGCCA GAACTCGTCG TCAACCTCGA
34141  CCGAGATGTT CGTCGTCCAG TGCTCGCCCG TGCTCGCCTT GATGTTGATG AACTTGTCGA
34201  TCTGGTAGTC GTCCCAGTGC ATCATCGACA TCCGCGCCGA CCGGCGCACA CCGCCGGCCA
34261  CAACACACTG AGCGATGGCG TGGTCGACCT CCATCGCGGC GATGCCGTCG AGCGTGATCC
34321  CTGCGTACTC CGAGAAGATG TTGGCGACCT TCTGCAGCAT CACAGCGAAC GGCAGCGGGC
34381  CGCTGGCCAC TCCACCGAAC GTCTTGAGCT TGGCCCCTTG CGGCCGGATG CGGCTCACGT
34441  CGTACACCCG CTGGTAGTGG ACCGTGCCGG GTCGGTAGTG CGTGTCGATC AGATCGACCA
34501  GCGCAGCAGC CCAGCCCTCT CGTGAGTCCT CGATGGCGTA GGCACCGGCC CAGTCGTGGC
34561  TGTAGTGCTC CGACAGAATG CCTACATCCT TCATCGCCTG GTAGTCGACA TGCTCTGGAT
34621  CACAGACGAT CTCGACCCGC AGGGGGTTTA CGACCTCGGG GTAGCCTTCG AGGTAGTGGT
34681  TCGAGTAGTT CGCCCCGACT CCCCCGCCCT CCATCAGGCG CATGAACGTG AACTGGAAGT
34741  GGTCCGAGAT CTTCTCGGGC CAGCCAGCTA CCCAGCAGTT GAAGAGGTGC TGCGCGTTCT
34801  TGACCCCCGA GGCCCACAGA TGCCGACCTG CCGGCAGCAC CTTGAACTTG GTCATCAGAC
34861  GAACGAGATC TTCTCGCTCT CCTTCCAACA TATGTCGCCG GTCGACAAGA GCAAGATTGC
34921  CGTCCACGAC CCTCTCGACC GTTTCCGGCC AGGTTTCCTT CGAGCCGTCA GGCTTGGTCC
34981  TGGCGTAGGT TCGGTTGTAA ACGAGTTCAC CGGTTGGTCC CCAAGGGATT TCGTCAGTCA
35041  ACTACTTCCT CTCAGTCAGT TCGTATCGCT TGAAATAGGC GTCGGCAGAG TCGCCGCCAG
35101  AGAACGAGAC CCCGTACTCG ACCGGGCCTG CACCACGCAC CTCGCAGGTA ACGACGCCCT
35161  TCCTTCCCCG GAACATCGGC CAGGTTCCCT TGGAGGGGTG CTTGGTCTCG TCCCGCTGGA
35221  CGATGACCTT GGTGCCCTTC TTCATGCCGA CTTCCGTTCT CCGTAGCCGG GAGTGAAGCA
35281  ACCCCCGACG TACAGCTCGA GATCTTCTTG CGACCAGTTC TCCAGTCGCA TCGGCGGCTG
35341  GTGCGGGAAC AGCTCCGGGA ACACCTCGGC CCGGTACAGC TCCGAACCGG GCATCCCGTT
35401  GAACGTCGGA TCAAGAATGT TGTGCATGGC ACCTCCCTCC CAAGAACTCG GAGATCGGCG
35461  GCTCGTAGAG GTAGCCATCG CGCAGCTCGG GGTTCTCGAT GAGCATGATC GCGATGTTCG
35521  CTGTGGGGTC AGAGTGCCCA TCCCCCTGCG ACTTTCGGAT GTCTGGGAAG ATAGCGTGCT
35581  TGCTGCCCGG ACCATCCTTG ACGATGACCT TGCCCTTGTC GTCCTTCTCC ACGCCAGCCG
```

35641 TGATCGCGAT GATGTTGACG TGCTCGGTCA GCGACTTGTG AGCGCGGAAC AACCGGTTCT
35701 GCCCGCTCTT ATCCTTCGGG GAGATCCCGT CGGTGTAGCG GCTCCTGATC GCCTCTGCAT
35761 AGCCCCCGTT CTGAGCGTCC AGAGCCTTCA TCGCCAGCGG GAGGATGTCG ACCAGGTACC
35821 GATTGGTCGA CTCCCCCTGC AGAGCCTCTT TGACGTTCTC GGACGAGTAG TGGCTGCGCT
35881 CCTGGAACAA GTCGCGGGCC TTGGCCGCTC CCGACAGGAT GTTGCGAACC TGATTGCGTA
35941 CGTAGTGAAC TGCCTCACCA CGGTGCAAGC TCTCCAGCGT CTTCTGGATG TACGGGCTCT
36001 CGAGGTACCA GACCCACAGC TCTTGGATGA TCTCCTCGGC TGTCAGGTTG GTCTCCCAAC
36061 CGATCAGCGC CTTCCGGGTG GCCCTGCTGA ACAGCTTGCT GATGTCGTCG GTCAAGGCAT
36121 CACCTTTCGT AGGTACTCCT CCCGGTCCAA TCGGCGGTCG AGGTGTCGAG TGACCTCCTC
36181 CGCGAAGACC TCGCGGACTT CGCTGGAGGT GATCTGGCGC GAACGTGCGT TCTTGTGCAG
36241 GTACGGCAGC TTGGTGGCTG TCAAGTTCTA GACCTCCCAG ACTCGGCCGT CGACCGAGAA
36301 CCGGCCTCCG ACAATCGGAA CAAGCTCAGG CTTGACGTGC TGGCCGTCGA CCGTCAGCAG
36361 AGCAAAACCA CTCTGCCAGT TGGCTGTTGC ACCCTTGAGG TACTGAGCTA GCTTCATGTT
36421 CATCAGGTTG CCGACCTCCA TCGACCACAG CACCTTCTGG TTGCCGCCGT AGCCCAGCGT
36481 GTGTGGCTTG ATGCCCTGGC GGTGGGTGTG TCCGATGATC ACCGACGTGC CGAACCGCAT
36541 CATCGCGTTG TACGCGGTGT CAGCGGACTT CTGCGTCACC CGGACCCCAC CACGGTGGCC
36601 GTGGGTGGAG ATCCAGCCTG GAGCGATCTT GTAGAACTCA GGCAGCACGT CAACACCGAA
36661 CCCGTCGAAG TCCAGCAGGT TCTGGAACTG GAACGAGCTG ACGTACTCGA CCAGCGCCGG
36721 GGCGAACTGG TGCAGGTAGT CGACTGGCCG GCGGTCGTGG TTGCCCTCGT GGACACCAAC
36781 CGGGCCGTCG TAGACCTGGC GCAGCGGCTC CAGGAACCGC CGCTTGCACT GCTCGGAGTC
36841 GGGCTTGATC CGCTGAGCGA ACTCTTCCTT GGTGCCCTTG GTCCACCGAG ACGGGCTCGG
36901 GTAGTCCATC AGGTCACCGA TGTGGACGAC CTCGTCAGGC TGGGTGTCCC CGATGTAGCC
36961 GATGACCGCC TTCAACTGCT TGCGATCATC GAACGGAATC TGGGTGTCCG AGATGACGAC
37021 GATGCGCTTG CTCACTCAGC GACCTCGGTG AAGGGGCCCC GCATACGTTC CTCGTGGGAG
37081 CTGGCGTTGC CTCCTGACCA GCGTCGCTTG CCCACCTTGG TGTGGTGCAA CCCGTTGGGG
37141 TAGTAGATCC ACTTCACTCC TGTGGCGTTG GTGACGGTCT TCACATCGGC AGGAACGTCC
37201 AGCAAGGTGT CCCACTGGCG AGGCCCCTTG GGATACCGCT CGTCCTCGGG GAGCTGCATC
37261 TTCTCCAGAA CGCCTGCGTA ACCGGCGATG TCGACCACCG TGTCCTGGTG GTAGCCGTTC
37321 TCCATGAACC GGGCGATCTT CAGCAGGATC ATCATGACGG CCACGTCCTC CGGGGTGAAC
37381 TCGACGCCGC GCTTGTACGC GCCCCACAGG GTCGCGATGC GTTCGTGGTT CTCCTTGGCG
37441 TCCCCGTAGT CCTGGGCTCG CTGTCCGTTG ATGATCTCTT CGGCGGTGGT CAGAATGCTC
37501 ACAGTCCAGT CTCCGATGCG GTGTAGTAGT CGATCAGCTC ATCGAGCTGG TCCGGTTGAT
37561 AGCCGAGGAT CGGCTTGTGG GTGTCAGTGA CGACGACGGG AACCGACATC GCGTTGAGCA
37621 CCTTGGTGAC GTAGTCGTAC GCCTCCGAGT TGGCCGTGAC ATCGACTGCG TCGAAGTCGA
37681 TCCCGGCAGC CGTCAGCTTG TCTTTGACTC GCTCGCATGG CTTGCAGCCG GGACGGGTGT
37741 ACACCGTGAC CGGCGCGAAC AGCGTTCTCA CGTGAGCACC ATCCCAGTCG ATGTATCGGT
37801 CTCCATACAT CAGATCCTTT CCAGCAGAGC AGCTTTGCCC TGCGATGTGA CTAGTGAGTT
37861 GACATCCTCG CCTTCTGGCA TCGGGATGAT TCGGGCGTTC GGCAGCGTCT TCGCCACCGA
37921 CCGGGCGAAC TCCATACCGG CGTCGTCGCC GTCGGCCAGG ATGTTCACGT TGCGGTAGCC
37981 CAGGAACAGC TCTCGGAAGT ACGGCTTCCA CTTCTGGGCT CCGCTGAGCC CCACCGTCGG
38041 CAGCCCACAC AGCTCGGCGG TGATCGTGTC GAGTTCTCCC TCGCAGATCG CCATGTCCTT
38101 GCTGTATTTG GTCAGCGCGT AGGTGTTGTA GAGCCGGTCC TTCTCCCCTG GCATCGACAG
38161 GTACTTCGGT GTGCCACCGT CGATTCGGCG ATACCGGATC GCAGCTACCG TCCAGTGACG
38221 CCAGGGCGAC CACCGCATAT ACGGAATCGC CAGGCAGCCC CGGTACATCT CATGTCCAGG
38281 GAGTGGGTCG TCCACGAATC CCAGACCGAA CCGGCTTAGT TCCGCTCGGC CGGCCAGCCC
38341 GCGACTCGCC ATATACTCGT CGGCTGGGCT TCCGGGCAGG CTTTCTCTGT ACCGGGACGT
38401 TGCCTCCCAC AGATAGGTTC TCTGCGATTC GCTTAGCCTC TGCAAATGTC ACCTCCTCTT
38461 CGTGACGAAT GATCGAGATC ACGTCTCCAC GGACCCCGCA GGCCATGCAG TTGTAGCCCT
38521 GTAGGTCGTA ACTGACTGCG GCAGACGGCG TTTCGTCGCC GTGGAAGGGG CACAGGCACT
38581 TGTTCCACTC GTGGTGGTCA GGTGGTGGTT CCCAATCCGG GTGGTAGCGA AGAATCGCCC
38641 TCGCGATGGG CGAGTCGTTC ATTCGTCCTC GTCAAGCTCC TCGGGAGAGA GCCCTTCGAA
38701 GATCCCGTTC AGGACGGCGG CGAAGCCCTC GCCGGTCTCC GCTGCGTCGA GCATCTCTGC
38761 AATCGTCTTT GCCATGTTTC CTCCTGGTGG ATGTCAAGTT CGAGACAGCT TGTCAGCCTC
38821 GACTGGAGCG ATGCGCTCCC CGATGACTTG GACGGCCGGC GGGTTCAGCA GGTACTCGAT
38881 GGCCCGTTTG AAGAACTCGA TGCAGTCCCT CGCCCAGCCC AGCGTGTACT TGTTGCACAT
38941 CGTGCAGAGC AACCCTCGGA CGATGCCTGT CTTGTGATCG TGGTCGACCG ACAGGCGCTT
39001 CTTCTTACCG TTGGCTCGCT GGCAGATGTA GCACCGACCA CCTTGGAACT CGTAGATCTG
39061 CCAATACTCA TCGCCGGTGA TGCCGTAGGT GGCCAGGATC CGGGTCTCCC AGCTCGTAGA
39121 GCTGCGAGCC GTCCTGAACT CTCGGTGATG AGTAGCGCAT CGTGGCCCTG GATACTTGGC
39181 GTCTCGCGTG AGCGGGAGCC CCTGTGCGAC ACAGTCTTTG CAAGGCTTCC GCTTGTGCTT

```
39241  ACGGTTCTGC ACCCGGTACC CCGGAGACCT CTTCGCCGCC CTCGGCACGC GCGTCCTCCT
39301  CCCGGTTCTC CATCACCATG CAGAACCACG ACAGCAGCCC TGCCAGGGAG ATGTAGAAGG
39361  CCACCAGAAC TTGGCCGCTC ACTTCACCAT TCCTCGAACC CACCAGCGAG ACAGCGCCTT
39421  ACGCCCTTTG TCGAGCGGGG TCAGCTCGCG CTCATCGTCC TCACCGAAGT CGAACTCGAT
39481  GCTGGCGATC TCGTAGCCGA GGATCTTGAA CGACACGTTC ATAGGCGGTC TCCGAAGTTG
39541  ATGACGGGAA TGCCGGCCCT TTCGGCCTCT CGCATGCAGT GCCGGGTGCC GACTGAGTTG
39601  CCGAGGGGGA ACGCCAGACA GATGTCCGCA CCGGCCCTGA CCATCTCGAT GTTGCGGAGG
39661  ATGCCAGCCC GCTTGCCGTA GCGTTCCCAG TCGGCTCGGT GCAGCTCGGG GAGCACGTCC
39721  CATCCCTCCT GCTTCATCCC CCAGGCCCAG CGGTCTGCGA TGTCGTCAGC GCCGCGAGCG
39781  CCGCCGTGGA CGACCGTGAG ACCGGAGAAG GACCGGTGGT ACTCAGTGGC CAACGCTTCC
39841  CAGACCGTGG TGCGGTCCTT CCAGATCCGA GATCCGGTGA TCAGTACTCG CCGCATCAGA
39901  TCGCCTCCCA CTGCAGGCCG TCGTGCGACG TGACCAGCTC CGCTTCGTAG ACGCCGTAGC
39961  GGGTGGCCAG GAACTGGATC ATCTGCGCCT GCTTGTACCC GAAGGGACAT TCGTGGACGC
40021  CGCTGATCGG GTATCTGACT CCGTATTTCA CTTGATCCAC CGCTTCGCGA TTCGGTCGAC
40081  GTTCTCCTCG GAGACGTTGC GGGCGAGGCC GGTGAACTCC TGGCCGTGGA CCTTGGTCTC
40141  GATCACGCGA GGCTTGCGGG GATCCGGGCT CTCCGGGTCG ATCCGCTTGT GGGTCCAGAC
40201  GGTCGGCTTC GTCTTGATCA GAGCGCCCAG CACCTGCTGG CGCAGTGGGT TGGTCTTGCG
40261  GGGCATAGCG TTTGGAGTGG TCATCTGGAT CCTTTCCTCG GTGGCTGTCA AGTCGGTGTG
40321  CGTAGTGAAG CCCCCCCAGG CATGCGCGCC CCGCCTGGGG AGAGTTGATC AGCGCAGTTC
40381  GATGTCGGGC AGGATCGCCT GCGGCTTGAA GTTGACCTGG TAGAAGTCGG TCGAGACGTT
40441  TGCGCCATCG ACCTGCTCCA TGAAGTAGGA GACGTTGTCC GACAGGCCCA GGAAGTGCTT
40501  CTTGATCCCG TCCTTGGTCT TGCAGGTCAC GTCGAGCTTC TTCGACGCGG TGTCCGCGTT
40561  GATTGAGCAC CGGCCCTGGA TCTCGAGCAG GTACTTGTCC GTGATCCCGT TGAAGAACAC
40621  GATCCGGCGA TTGATCTCGA AGTTGTCAGC GGCCTTGCTG ACGTTCTCCG ATGCGACGTC
40681  GGCGTCGGAG GTACACGCGG AGAGGCCCAG GATCGCCGAT CCGGCGATGA GTGCGGTGGC
40741  GATGATCTTC TTCATGTTCG CTACTTTCTG TTTGGTGGAT GTCAAGTTAG TGACCGAAGT
40801  CGTTGATCTG CATAGTGTCT CCGACGAACT CCAAGGAAGC GAAGTCTTGT CCCGACGGGT
40861  CCGACTTCCC CCCTCGGTTC TTGACCGTGG AGACGTTGAG CATGTCCGGG CCGAACCCGT
40921  CCGATACTCG GTGGAGAGTG AGGATCATCT CAGGAACACG CCCGATCTGA CCTTTGATGC
40981  CCGACAACGG GATCGGCTTG TCGCCGTCGT TGTGCGGGCC GGTGACGTGG TGGAGCCCGA
41041  CGACGCATGA GCCTGTCTCA CGGCCCATCT CGTGTAGGTA GTCCATCAGC GACTCCAGAC
41101  CCGAGAACGG GTCGTCTCCC TCGCTTGAAT CGGTGCGGAC GTTGGTGATG TTGTCCACGA
41161  CGATCAACGC TGGGAAGTCC TCGTACAGCG CGTCATACGC GGCCAGAGCG TTCTCGATCT
41221  CGTCCAACGA CGGTGATGCC TTGTAGTTGA ACCGGATCGG GATCTCGTCT AGTGAGTCAG
41281  CTACCGCGTC CTCGATGTTC TGCTCGCGAA CAGCCCGCGT AGCTCGTTCG AGCGACCATC
41341  CGCTGAGGAT GGACACCGAA CGGGAGAGCT GGGTGAACGC ATCAGAGTCG GCCGAGAAGT
41401  ACAACGTCGG CACCTTCGAC TTGAGCGCGT AGGCGAGGAC GAACGCCGAC TTCCCGGTGC
41461  CGGGGCCGGC GCAGACCAGG ACTAGCTGGC CTCGTCGGAG ATGTGTACCT TTCTGGTCAA
41521  GCGCGGCCCA GACCGGGGGT AGCGGATCCC CCGCCGACCC TCGGATGTAG AGCGATTGTC
41581  TAGGTGTGTA CACCTTCCTC CTCGTGGATG TGATTGACCA GGTCATAGAT CTCGTCGCGA
41641  GAGACCAGCC GGCCCCAGGC GTCGATCCCC ACGTGGATCT GTCTCCGGTG GATGTGTCGG
41701  GACAGGATCA TCGGCGAATG CGTGTGCCCG TGGATCAGGA TCTTGCCATC GTCACGGAGC
41761  CTCCACTGGG TGTGTCGGTC CTCGCTGGTG TGGTCCCCGA CGTATGGGAA GTGGCTCAGC
41821  AGAACATCTG TGTGCCCGCC AGCGTCCCCG TACAGCGGCA CCCGGATACG AGCTGCCGTC
41881  GACACATGCT CGAACACCAT CCAGTACGCA CCAACCAGCT TGTGAGCATC GCGGTTCATC
41941  GGGTGGGGCC CATCGTGGTT GCCCAGGATC AGCCGTTTGC GGCCTGGCCG ATCCGAGATC
42001  CACCCGAGGG CATGTATCTG CCCCTTGGTG GAGCCAGAGG AGATGTCACC TAGGATCCAG
42061  ACCGTGTCGT CCTTGCCGAC GACCGAGTCC CACGCCTTCG CCAGGGTGGC GTCGTGCTCT
42121  TCGACATCAT CCGCCAGGTT GCGGATCTCC ATCAGCCGCT TGTGTCCGAT GTGTAGATCG
42181  GACGTGAACC AGGTGTTGCT CATGGCTTCC TTTCAGAACG GCGGGCCGTA CAGCTCGATC
42241  ACCAGCGCGT GCAGCTCCTC TGCCGCGTCG TCACGCTCGA ATCCGCAGCA GGAATCGTGC
42301  CGGTCGAGGA TTGCGACGAT CTGGTCGTAG AGGCTGGGCC TCACTTCACC TTCTTCGGAT
42361  CGATCAAGGC GTCGTGAATC GGCCGACCGG CGCGAGCCGC GTGCGTCTCG GCGTCCAAGG
42421  CTCGCTGCAT CTGGTTCATC AGCCGGGTGC CGCGCAGCTT GAGGATCTTC ATGGTCGCCC
42481  GACCCTTGTA TCCAGCGCGG TGCATCCGTA GGACGCAGGC TGTCTCGTGC GGGGCTATAG
42541  GTGACCTCAG CGACGGGTGG TTTGGATCCC AGTTCGTCAT GTCTTCCTCT CGGTGGCTGT
42601  CAAGTTGGTC ACAGACCGAA CTCTTCCTGG TACTGCGGGA TGAAGTGGCC GGCCGTTCAT
42661  GTTCGGCTCG ATACCTCTCG CGTCACGAAC TCCTGCCCGT TCCATCTCCG ACCGTCCTCG
42721  AACTCGATCA CGATCTCTCG TCCGGGATGA CGCACGGCCT CCGCTTGGGC AAACCTGCGT
42781  GCAGCCTCTG GGGTCGGGAA CGGAAACTTC TGCGAGGCGT ACAGCTCCTG GTGCCACTTC
```

```
42841  GGCTTGTCAG GAATCGGCCC CATTTCCACG TACGTGTAAC CCGCGTCGGG GTCGAGTTCG
42901  AGCGTTTTCT TGTATTCCTT CGTGCCTGCC TTAGAGGGAA GGTGAGTATC GGTGGCTGTC
42961  AAGGTGACCT CACTTAAAAA CAGGGCAGCT GTAATTCACA TCACAGAAGC CGCATTTGTC
43021  AGGTTCAGGC AGAGGCTCGA AGTCACCAGC CTGGATCCGA GCCTCGACCT CATGGAACCT
43081  CTCGGTGATC CGCTCCCGCG TCCAATCGGT CAGGTCGTAG GGCGCAGTGG GCTTCGCCTT
43141  GATGCCCTTC TTCCCCGCCA TGAAGTAGTC GCCCGTCTTC GGAGCCTCCA CGTCATAGGT
43201  CATCGCGACC GCGAGCGCGT ACACGCCGAG CTGGAAGTCG TCACCCGGCG AGTTGCCGGT
43261  CTTGTAGTCC CGGACTCGAA GCTCACCGTT GACCACGACG ACCGCGTCGA TGAACCCTCG
43321  GACGCGGATG CCGTCCAGCT CGATGTTGAA CGGAAGCTCG ATGGCCGGCT TGGGCTGTTC
43381  ACACTCCTTG CAGTTGGTGT CTTTCCACGC CTCCGTAGAG CAGATCCCTC GCCCAGGGGT
43441  AGTCCAGATC TGCTGGCCCT TGTCCTTCCG CCACGCGATG AACTTCTCTA CCTGCTCCAG
43501  TCCAAGGTGG AACCGGCGCT CGATGTCACG CTCACCGTTG TACGGCCCGG ACCAAAACCA
43561  CCACTCGAAG TTCGGGGTTT CGTCGCACAG TGCTCCGATG TCCTTGGCGT ACTCCTCGCG
43621  GAAGATCTCT TGTGCCCGTT CGAGGCTCAT CTCGCGGCCC TCGGCCAGAG CCTTCTCGTA
43681  GACCTCAGCG ACGGTGTGAA ACGCGGTGCC CTGCGGCAAC CACGCCGCAG GACGAGCCCA
43741  TACCTTGTCG ATGCGAGCCA GCTTGTACGC CTGCGGGCAA CGTGTGTATT GGTTCAACTG
43801  GCTGACGCTT CGCAGCGGCA GCAATGTCTT GGTGTCTGTC ACGCAGCGGC CATCCTTCCC
43861  TTGCCTATCG TCTCGTTCAG CGCCCCGTCG ACAGCGACAC TGAGCAGTTT TGCGACCTCC
43921  GACATGTCAA TCGGATCCTT GGGGAATTGG TCAGCCTGAG TCATCCTGAG CACCATCCAC
43981  TCGGTGCCCT TGTCGCAGTG GATCATGGTC GGATCAAAGC GAGTTCCCCG TGCTACGTAC
44041  TCGACTTTGT TCGCGGAAAG AATCAAATTC GACACAGGCC GATAAAGTCG TGAGGTGTCT
44101  TTTACACGAG GACTGCGGTA GACGAGCAGA ACTGAGACTG GGTCTTCGTC CAGTTGGCCC
44161  TTCCACCACG CCTCACACCT CTGCGCGAAC AGCCACCCTG GATGATCGGC GATGACTTGC
44221  GGTGAGGTGT GGACGAGGTT GTCTGCGAAC AGCTTTGCGA GCCGAGTGAG GGGCACGGGG
44281  TTTCCTTTCG TTGCGCGGCC TGGGTTGGCT CACACAACCG GTCGTGACTT TTAGGGCTCC
44341  GAGAGAAGCT CCTCGATGTC GTCTGGCCAC GACCAGAGGA GTTCACCCTC GGCGGTGAGG
44401  TTGGTGTGCT CGTTCACCCG GATCAGGAGA TCGTCATCCT CGATGCCTCG GGGGACGTAC
44461  CTGAACCCGC CGCCGGCCAT ACCTTCGTAG GGCTCGATGG ATGGGTCGAA CTCGAGCACT
44521  AAGTCGTCGT CGCGGAGCAT CTTCCACCAC GACAATAGGC GCTTCTTCTT GTCTTCGGAC
44581  ATCGTGCGGA AGCTACCCAC TCGCATGTAC TCGCCGTGAT CCCGGAGCCT CTGAAAAGCC
44641  TTCGACTTAT CGTGAGGTTT CCGCGTGTCC CACGGCCAGT TCTGCTGGAC GATCTGCCTG
44701  GTGGTCAACC GTCCTCCGTA GGTCTTCTTG TGCCACGACA CCGCTTGTCG AGTCACGCCA
44761  TACAGCTCTG CGATTTCGGT CTGATTAAAC CCCTTCCTGC GAAGATCTTC GATCTCGCTG
44821  AGAGTGAGTG GTATTCGGCT AGGGGCCGGA ACCACTGCTT TGTGTTGGAT TTTGCCGCTC
44881  ATGTTTCCCT CCATGAGAAA GGTGCGTGCG TCTCCGCCGA TTACGGAGAC ATGTTGGTGC
44941  CTGTCAAGGA TACCCCTAAT TTAGTTGCGT CTGCGGAACC ATATTCAGTT GTGTTCCCCG
45001  ACGCCGTGGC CGTCTCCCAC TGGGCGTGGG ATCGACTGGC GTTACGCGGT CGTAAATGTA
45061  GCGGCCTGCC CCACTCGGTA GCAAACCTTG TGACAGGTAT CACTTAGGTC GCCTTCTGTT
45121  ACACGTTGAC CTCGGGTTTC ATCGTCACGA CTCTCCTTTC TTAGACAGCC TCAAGATCGT

45181  TACACCGGCT TGCGAAGATG TACCTTCGCC TTGAATCCGG CCCTTGCCAG CTCGAACTCG
45241  ACCACCTGGC GGGCGGTCTC CTTCAGGTCG GACTTCGCCG ACAGCGGCCC GACGAACCCG
45301  TAGCTCTTGA TGTACTCCTC GAGGTCGATG TCGACGTACA GCGTGACAGG GACCACCGAC
45361  AAGTCACACC TCCAATTCGT GGGGCTTGAT CTCGTTGGTC ACGTCGTAGT CGTTCAGCAG
45421  CGACTGGAAG TCGGAGTCTG TCAAGTCGTC CAACTCATCC TGCTCGAACG GCGCGGGCTC
45481  GTCATGCCAC GTCTTCCACT GGTCGTGGTC GGCGCGGAAC CACTTCCGCA GATCCTTGAT
45541  GGCCTCGTCC TCGGTGGCGA AGACGTAGGT CTCGAGCACG TCCTCGTACT CGACGGTCAG
45601  CGACCAGACG GTGATCTTCA CTCCCCGTTC ACCTCCGCTT TGTAGTTCAT CTCGGCGGTC
45661  TCCTCCTAGT TGGGTAGCAG TCGGTTGTAC TCGTCGTGGC TGATCTCGCC AACGATGAAC
45721  TGGCGCATCA GATTTGCGAC CGAAGCCGCG TCCATCCCTT CGGGAATGGG CTTGGCGTGG
45781  CCGAACTGCC AGTCTCGTGA GCGCCAGCGG AACCAGAGTT GGACCTTGTC CAGTGAGGTC
45841  AGGTGCAGGC ACTGAAACGT CATGCCTCCG AACGGGAACT CCATCACACC TCCTGTTTGA
45901  CCTTGACGGT GTGGCCTGTC ATTACTTCGT GGATTCGGAT GCTGGTGCCG AACGTCTTTC
45961  GCGTCTCGGC CTTGAACTCG GTGGAGCACC CCGAGCACTT CGCTTTGAAT CGCACTAGCA
46021  GTACCAACGC TTTCTGCAGA ATCGGGACTT GCCGCCGTCC CGGTTGTCGT TGTCCCGGCG
46081  GGCTTCGCCC TTCGGTGATT CGTCACATGA CGGAAGCTCG CCATGCTTGA TGTGCCATGC
46141  GTCGTCGGCG ACTTTTCCGC CGTGCTCGGC GATGTGCGCT GCGCTCCGGT ACTCACAGAG
46201  CGGGGAAGCC GATGCCTCGG CGATGATCCC AGGCAGGTTG CCTAGAACCA CCGCCAAGCA
46261  CATCAGCAGA ACGACGTGCC ACGCCTTCAT CAGCCCGCCA GCGCGTGGTT CATCGCCGCG
46321  TTGCGGCCGT CGCGCTGACC GTGGGCATAG CCGCTGAGGT CGTACCGGGT CCGAGGCTTG
46381  ACGTTCTTGG TGCGAGGATG CGCCTGGCGC AGAGCCAGCG CAGCTCGTTC CTTGTCGCCT
```

```
46441  CGGTAGAGCA CCAACGCTCC CCCGCCGGCC GATTCCACGG CCTTGTTCTC CTCGGCGGTC
46501  AGGCGTTCCT TGACGGCCTG GGCGAAGCCT GCGATCCACG ACCGGCGGTA GCTCTTGAGC
46561  TGGCCAGCGG TGCTCTTCGG CTTGTACTCC CCGGTGTTGT AGTCGTACTT GTACCGAGGC
46621  TCGAAAGCCT GCTCCGGGCG GACATTCTCA ACCAGGCGCA TCATCTGCGG CTGCATGATC
46681  GACCAGAGGA ATTGGAGCCT CTCGATGTGG CGGGGCACGC CGTAGACGTA GATCCGCTGA
46741  CCGCCCGTGA GGCTGGCGTA CACCGTCTTG CAGTGCAGGG CCTGAGCCAT GCCGTGCAGC
46801  AACAACGCTT GTGCGGCAAC GTACTTGCCG GTGACGTAGG TGACCCACTG GATGGCGTCG
46861  GGCAGGTCGG TGGTGTCCAA CCCTTGCTTG CTCGCCTCGA CCTGGGCCAT CTCCAGCCCG
46921  TACTTGGCCA TCAGCTCGAA CGCTTTCGCC TGGAACACAG CCTCTTCCGG CGTACCGGCC
46981  ACGTCTTCGG CCTGGCGCAG CAGCTTGGCG ACCTTGTCCT GCATCTTCTT CGTCTTGCCG
47041  TCGATCATGG TCAGTACTCC TTCTTCCAGT TGTTCCGGTT GCCCTTGCCG GGGCGCTTCA
47101  TCTCTCGCTT GCGGTTACGG TGCGGCTGCG CCGCGTTGGA GAGACGCAAC TCGAGCCGTG
47161  CCTTGAGCTG GTCGCTCATC TTCTTCACCT CTTCTGGTTC AGCGGATCTG GTCGACGTGG
47221  ATGCAGCCGA CGCGGTCTGG CCCGAACTCG GGAGCGAAGC CCAAGACTTC GTCCTCCTCG
47281  CATGGGAACG CTCGCTGGTC GAACGTGATT GGGTCGGCCG AAGCCTCGTA TGGATCGGCC
47341  AAGGCCATCG CTCCGACCGC TGTAGCGAAT GCAACGACGA CGGTGATCAG GTGCTTCTTC
47401  ACTCTTCTTC CCTCCACTTT TGGTCTGCGA GAAGCCTTCT GGCGATCTCG ATAGGTTCGA
47461  TCTCAGGAGT CACTCATCGC CCTCCAAGAT CTTCAGGTTG GCCAGCAGTG CATTGGCCAC
47521  AGCTCCGATG TGGCCACCGC CCTTACCTCC ACGGCGGGAG TACTCGCGGT TCGCGGCCTG
47581  CATGAAGTGG AACCTCGGTG AGCCGTCCTC GTGAACCCAC GAGGCTTTCT CGGCGGGCAG
47641  AGCCCGGTTC ATCTCCACCG ACATCGTGAC GATGATGTGG TCCCTCTGGA GCCGAGCCTC
47701  GGTCTCGGCG TAGTGGGCAG CTTGGATTAC TGCGCCTCGT GTGGTCATGT CTTCTCCTTC
47761  GGTAGATGTC AAGCTGTCGT CACCACTCTT CGACCGGTAT CGGTTTGTCA CAGCCAGCAA
47821  GGATCGCGGC GTTGCTGCGG TGATGCCCGT CCCACAGCGT CTTTCGGTCC CTCGAAACCT
47881  CGAGGGGTTC GAACGGCCAC TCGTTCGATG AGTTGAGGAT GTCCACGACT TCGTGGACCT
47941  TGGCCCAGAA CTTGCCGGTC ACGCCTCCCT GGTAGTTGTA GCGGGGCGTG GTCTGGTAGA
48001  ACTCTTCGAG CACTGGTCCG CTGTCGGCGA CGGTGCAGTC GACACCAGCG CAGGACATGC
48061  AGTCGCTGGC GCGGAGCTGG GCAACTTCAT CGGTGGTCAT GAACGCCGTG GTCACATCGA
48121  GCCTTTCAGG TGTATGTCAA GCGGCGCGGA CGCCGGAATC GGAGAGGTAG ACGCGGTCAG

48181  CTCCCAGGAA CGGAGCCTGT GTGTTGGCGT GGACGAACGT GTCGTTCTCG TAGGGGTTGT
48241  AGGCGATCTT CGATCCCACG AAGTCTTGCG GGAGAAGCGA GATCAGCTCG CCTACGATGC
48301  CAGCGTGGAC CACCTTGCGG CGCTCGCGCC GTACCTTGTC GCGGCCGGCC GGCCGAACCA
48361  CACCCTTGGC GTGGGCCAGC AGGACGTGGC CGCTGCGGTG GATGACTCGA CCCTTGAAGT
48421  CTCCCTCCAA GGCTTGCACC GAGTACCACG GCTTGCCCTC GCGGTGCGTG CGGTGCAGGT
48481  TCTTGTAGAC GAAGACTCGG ATCGGCTTGG GAGTCATGAG ACCTCCAGTG TGCGAACGGC
48541  CTTGTAGGCA CTGATGAGTG ACGCCCCCGA CAGCTCGTTA CCGTGCAGGT GATACCTGTA
48601  TTTCAGATAC ACGGCTTGGT CGACCGGCTT GTACTCGACC GAAGTGACCT CGACAACCAT
48661  CCCGTCGATG ATCGCGAAGT CTCCAGCGCG GAGATGGGTG GGGAATTTGA TCTCGGTGTT
48721  GACTACGGTC ACAGCTTCGA AACCTCCCAG GTACCAACGA ACTTGCCGTT GCGCTTGATG
48781  TATCCGCTCT CACCGGGCTC GTACCAATCG ACCTCGAACC CGTAGCGGGC GGCGCAAGCC
48841  TCGAGGTGGT CGAGCAGGAC GCGGCGACCG GACGCGGTAG CTTCTCCGGT CAGCCCGCTG
48901  TCGTTCTTGC GGACGATGAG CTTGAACACT TGGTGCCTAC CCTTCTGCGA TGTCTCGGGA
48961  GATCTCGGCG AAGACTTTCT TTGCCCACGC CACGCCGTCC CAGGTGATGT CGAACAGTGC
49021  CTCGTAGAAC TGGTCTCGCA AGGCTTCGTT GCCGTCGGCC AGCGTTGTGA CGAGCCGGTC
49081  GATGCGGTCC TCGTGGAACT TGTAGACCGA GTGGTTGTAC GGCTCAGCCA TATTGGCGTT
49141  GGCTCGTTTC ACGTTCTCAA CCACGATGGC TTCGAATAGG TGGTTAACCA GCTCCTCGGT
49201  CATGTTCTAT CTCTCCTCAG TAGTCGCTGT GCTGGGTCTC GAAGCCTTCG AGGTCACCGA
49261  CCTCGTCGTC GTACGCGCTC GGGTTGCCGC GCCAGTCGTC GCGGAGCCTT TGACCGCTGG
49321  CGTTGTAGCA GGCACCACAG TTCGGGCAGT CCACATCGCT CTGGCCGTAG TAGCGGCAAA
49381  CCTCGCCGCC GCAGCGTTGG CAGTCCCACG CGCTGTAACC AGGGATCAGG AAACCTTGGT
49441  CGTCGGTCTG ATCAGGGATG CGTCGGAAGT TCTTGGCAGG CATAGCTACT CCTCATAGAA
49501  ACTCGTGGTT GATGGCTCGG TGGGCAGCCT CGCGGAAGGT CAGCCCGTCG TCGTACGCGT
49561  CCCGGTACGT CCAGTCCGCG ATGTCTTGGT AACCAAGACC AAAGGTCTCG GTCATGTAGC
49621  CGTCCAGCGC GGCCATCCAG GTCTCGAAGC TCATGTCTTC CCTCACTTCT TTGTGGTCGA
49681  GAACAGCACG TTCCTGCGGC CGTTGACGCA CAGACCGCAA CGGGCACAAG CCGATCCCTT
49741  GTCGTTGATC AGGTCGATGG CTTTGTTGTT CTCCGGGCAG CGCACCGCCG TCGGAAACTC
49801  GGCCTTGCCT TTGGCGAACG TGGTGTCGAC GTAGGCGATG TTGATGCCCT TGTCTTCCAA
49861  GAAGCGCGCC ACGTCGATGT TGTCCGGGTC TGCGCTGAAG TACAGCGCCA GGTTGTCGAG
49921  CCTCTGCGAG TGCAGGTAGA CAGCCGCCGT CTGAACCCTT GTGTAGGCCC AGAACTGGAC
49981  ATCCGGGTTG TCGCGGATGA CTCGACCCCA AGCGGCCACA TAGGTGGGGC TGAAGAAGTC
```

```
50041 TCCATCCCAG TGGATGCGGA ACAGCTTCGG AGCCTTGCGA CGGTCGCAAT CCTTGACGAA
50101 CTCGGCGACC ATCTCGGACA GCAGCGTCAC GGTGTCTGTC AAGTCAGCGT CACGCAACAG
50161 TTCCCAGTTG TGCAGCAGGA CCGAGCTGAC AGCCTTGCGA ACTTTCTCCA GCTTGCCGGC
50221 GTAGCACACC TTGGCACAGA AGGCCGTCGC GTCCGGGCAG GAGAAGCCTT GACCGGAGGG
50281 CAGGCCGATG CTGTTGGCGA TACCTACGGT GGCGTTGCCG CCCTTGGTGA CGTGGACGTA
50341 GTTGGTGACC TTGCGGTCGT TCGAACGCTT CAGCTTGGCC ATACCTAGCC TTCCTTCGGT
50401 GGCTGTCAAG TTGTTGGATA CAAAGCGCCC CGAGAGGGAG TCGAACCCTC ACACCGCGAA
50461 CCGTCGCGGG GCCACCGTGC CTAGTCGATA GAGGTCACTC GACTCTCGTG GACGTAGACC
50521 ACGGTGTTGC CTACGTTCAC CGCGTAGTAC AGGCCATCGG CACCTCGTAG CTTGTGCCGA
50581 ACCGTGCCCG ACGTGGCCGT CATGTCTTCG CCCCAGTCGG CGTTAGGTGC CCAGGTGACT
50641 CGCATGGTGA TCCCTTCAGT AGTCGGTGGC TGTCAAGTCA GCGGATACGG ACGTACCCGT
50701 TGCCTCGAGC GACGTAGATC TTGCCGTCGA TGTAAACGCG CTGCTGCTGG TTCATAATCC
50761 TATTCCTTTC GGTGGCTGTC AAGTCTCAGG CCCAGCGACG AGTCGTCGGC CGGGGGCGGC
50821 GCACCTTGGG CGCGTTGGCT CGCGGTGCCT TACGGATGGC GGTGCCTACC GTGATCTCTT
50881 CCAACTGGCG TTCAGCCAGG CCGACAGGCC GGGCGTCACC GGGCAGTTCG ATCTTGTAAT
50941 CGAAGTCAGT CCACCCCTTC AGACCCTTCT CCAGCTCGCG ATCCAACAGA CGCGGAGCCG
51001 ACAGCTCAGG CGCAACAAAC GGTGTCTTGA CGCTCTCGCG GGCAGTAACC CGAACCTCAC
50161 GGTGCTCAGC GAAGACTGGC ATAGTTCACC CCTTTGGTGG ATGTCAAGCC TGAGCACCAA
51121 AGCTCAGGCG TAGTGGGTAG TCGGGAATCG AACCCGATAG CTTCATAGCC ACGTTCTACG

51181 GCTCAGCCAT AGCTCAGCGA TCATTCCATC GCGCCAAGAG CTACCCTCCC GAATGCCGAA
51241 CCAAAGCTCA GCATTCGTAA GTGTGTATTC TCCCCGTGGC TCAGACAGTA TCTATCAGAA
51301 CCTAACCACA GGTCTACATT TAGTTATCCG CAGTGCTCGC ACTTTAACGG CATCGAGCTT
51361 CCGCCGACCC TCAGTCCTCT GGCAGCGAAC TAAAGGTTTG AGTCGGGCTG CGGCCCTTCT
51421 CGGTCTTGCG TGATTCTCAC TCTACCGGAT GTTTCGGTGG CTGTCAAGCG GGCCGTTTTG
51481 GTGTTGCAAC GATGCCCTCG TTTAGCGCCG CTGGCGTAAT GCGCTACCCG CCTGATCTCA
51541 CCGGTCCAAG TTGGTGATGC TTGCAGCTTA CCCGATAACC GGGTGGCTGT CAAACCGGAG
51601 AATCTTGCCG CCGGATTTTC ACCGGCACCG GCACGATCCT CTCGGATCCG CCTACCGCCT
51661 TGCTGCTGCG GTCCACACAG AATGCACTAC TGGCCGGGTG GCTGTCAAGC CCTAATCGCA
51721 AATTGGTGCC CTAGCTGCAG ATATGGCGCG TTCTCGGTGG CTGTAAAGGG CACTACGTGC
51781 CGCTATCCGC TGGTCACGCT GGACAGTCCC GGCAGCCCGT GCCGCGCATA GGCTGCTCAC
51841 TACGTGCCCG GTATCGGCGT TGTCGTGCCG CTGTCGTGGT CGTCGCCCCG TCGCTGTCGC
51901 TGGTCTCGGT GGCATCGCTT GACAGTCGCC CCGCTATCCC CCGTTGCCGC TGGTCAGACG
51961 CTAATCCGCT TATTTCGCAT AGGCTGCTCA CTATCGCATC GGTATGCGTA TGCGCTGGTC
52021 ACATATGCGT GTGGTGGTGG TGTGGTGTGC GTGTGTTTGC GCTGGTCAGC CGTGTGCGTA
52081 CCGTATCCGC ACACTGTGCT TGTGCGTTTG CTGTGTGTCG AGGCCGGCTC TCGCATCGTC
52141 GCATGTCAGC GCGGGTATGG GCGTGTATCG CACGCTTTGC TAGCCGCGTG CCGCGGCGCT
52201 CTCGCATCGC ATCGAGTGTT TGCTGTGTCT CTCATCGTCG CAGGTCAGAA GGGGTAGGGG
52261 GGTTCCCCCT AGGGGTCGGT CCTTGACCGG TCGGTTAT
```

It is known that during the establishment of lysogeny, the L5 genome becomes integrated into the mycobacterial chromosome and the attachment site (attP). Integration-proficient plasmid vectors have been constructed which efficiently transform both fast-growing and slow-growing mycobacteria through stable integration of the plasmid sequences into the bacterial chromosomal attachment site (attB).

Because the L5 sequence is now known, and because L5 has been previously characterized, the use of transcriptional promoters with this mycobacteriophage may be evaluated efficiently, and host synthesis inhibition may also be evaluated efficiently.

Figure 1 represents the genome organization of the entire L5 genome. DNA analysis has indicated that the L5 genome is organized into a right and left arm with the attachment site and integrase at the center of the genome. The integration functions have been successfully employed to construct integration-proficient vectors for mycobacteria.

Part of the L5 genome is not essential for mycobacteriophage growth. It has been demonstrated that all or most of the gene 62-61-60 can be deleted without affecting the cycle of the L5 phage. Therefore, there is a suitable region in the L5 mycobacteriophage for the insertion of reporter

genes. It is critical that reporter genes be inserted into non-essential regions of the mycobacteirophage. Otherwise, the mycobacteriophage will be unable to survive and replicate.

The L5 mycobacteriophage may have introduced therein promoter gene 62 fused to reporter gene lacZ, and this reporter mycobacteriophage will be capable of rapid diagnosis of mycobacterial infection and accurate assessment of mycobacterial strain drug susceptibilities.

Another mycobacteriophage which may be successfully used to produce the reporter mycobaceriophages is the mycobacteriophage TM4. TM4 has been used to construct a first generation reporter mycobaceriophage, and has the ability to discriminate between M. tuberculosis and BCG. A shuttle plasmid may be employed with TM4, and may be useful in the construction of recombinant and other mycobacteriophages. Unlike L5, which is a broad host-range mycobacteriophage, TM4 is a species-specific mycobacteriophage. However, TM4 is not as well characterized as the L5 mycobacteriophage, and therefore it is more difficult to analyze its functions.

DS6A is a mycobaceriophage that has been found to be specific for the M. tuberculosis complex of mycobacteria. It has been shown to inf ct both

M. tuberculosis and BCG. It has been demonstrated that DS6A can infect over 3,000 different types of M. tuberculosis strains. Current efforts are under way to develop DS6A shuttle phasmids containing Firefly luciferase genes as the reporter molecule. It is possible that a combination of different mycobacteriophages may be needed to increase specificity and then increase the ability to distinguish drug susceptibilities. DS6A grows on BCG and M. tuberculosis, but does not grow on M. smegmatis.

In anticipation of the need for a diverse set of mycobacteriophages that can effect a broad or limited range of mycobacterial cells, a total of more than 50 unique mycobacteriophages have been collected and isolated by the inventors. 21 new mycobacteriophages have been isolated from soil samples from India, France, England, Israel, Tunisia, Carville, LA and New York. In addition, another 30 mycobacteriophages from both the Centers for Disease Control in Atlanta and the World Health Organization Phage Reference Laboratory in Amsterdam were collected. The characterization of the nucleic acid content of the phage particles of 30 of these mycobacteriophages have revealed that all of the mycobacteriophages contain double stranded DNA whose genome sizes range from 45 to 100kb as sized on pulsed field gels. Restriction analysis has shown that all

of these mycobacteriophages are different, except that one of the mycobacteriophages from France had a considerable similarity to the L5 mycobacteriophage, which was originally isolated in Japan. The host range of the mycobacteriophages varies greatly, some being able to infect only M. smegmatis and others being able to infect M. smegmatis, BCG and M. tuberculosis, but not M. avium. These mycobacteriophages may be developed into reporter mycobacteriophages and cosmid cloning systems, and may provide a source of useful transcriptional translation initiating sequences, transcriptional terminators, or host-range specificity genes.

In addition, the choice of reporter gene and its method of expression are critical. It is necessary to choose a reporter gene whose product would not normally be found in clinical samples, but whose product is also easily detectable.

Luciferase reporter genes have been used in many diversified biological systems, including E. coli, cyanobacteria, phytopathogenic bacteria and Bacillus. The presence of luciferase reporter genes can be detected by the emission of photons in the presence of a substrate, such as luciferin or decanal. Luciferin and decanal can permeate mycobacteria, and ther by allow for th detection of gene products, such as photons. Sinc one molecule of

the luciferase gene product can yield 0.85 photons of light, it is the most sensitive biological reporter molecule known. The preferred reporter genes of this invention are luciferase reporter genes, such as the Firefly lux gene (FFlux), the Vibrio fischeri lux genes and the Xenorhabdus luminescens lux genes, as well as the E. coli ß-galactosidase (lacZ) genes. Luciferase genes, especially the Firetly lux gene, geneate a high amount of luminescence activity. They generate photons, the detection of which is simple and sensitive, using commercially available luminometers that can detect 100-1000 molecules of luciferase with a linear relationship to enzyme concentration. In addition, it is unlikely that clinical samples will contain significant levels of endogenous luciferase activity.

In choosing transcriptional promoters to be introduced into the mycobacteriophages, it is desirable to use strong promoters since this will increase the sensitivity of the system. In addition, it is important that the promoter be active following mycobacteriophage infection. The best promoter candidates currently available are the BCG hsp60 promoter and the L5 gene 62 promoter, which are of comparable strength. The hsp60 promoter gives good levels of luciferase expression from plasmid recombinants, but lower levels of luciferase

xpression where the mycobacteriophage is TM4. It is possible that the reason for this is that the hsp60 promoter is shut off by the TM4 enzymes following infection, thus producing only a modest level of luciferase. The gene 62 promoter may behave in a similar manner with the TM4 phage since the gene 62 product is a good candidate for the L5 repressor and is expressed at high levels in the absence of other mycobacteriophage functions. Knowing the sequence of the mycobacteriophage used will help in identifying, characterizing and cloning the appropriate promoter to be used in the reporter mycobacteriophages of this invention.

There are several methods which can be utilized to introduce the reporter genes and transcriptional promoters into mycobacterial species-specific mycobacteriophages. One method is the utilization of shuttle phasmids. When utilizing shuttle phasmid technology, it is necessary to know the sequence of the mycobacteriophage so that the reporter genes are inserted into non-essential regions of the mycobacteriophage. Insertion of reporter genes into non-essential regions permits the mycobacteriophage to survive and replicate. In order to use the shuttle phasmid methodology, it is necessary to first generate a cosmid library of large double-stranded recombinant DNA fragments of

mycobacteriophage. This can be done using cosmid cloning in E. coli. Next, the cosmid library is introduced into the mycobacteria of interest to select for cosmids which have been inserted into non-essential regions of the mycobacteriophage. The shuttle phasmids, which consist of the E. coli cosmid, the reporter genes and mycobacteriophage promoters, may then be characterized. Shuttle phasmids can be propagated in E. coli as plasmids, and propagated in mycobacteria as mycobacteriophages.

A second method of introducing the reporter genes and transcriptional promoters into mycobacteriophages is by homologous recombination or PCR. First, non-essential regions of a mycobacteriophage must be determined. Again, in order to do this, it is necessary to know the sequence of the mycobacteriophage. Consequently, L5 is an ideal phage to use with this method as its genome has already been sequenced and characterized by the inventors. Next, plasmids are constructed wherein reporter genes hooked to transcriptional promoters are flanked by mycobacteriophage non-essential region sequences in mycobacterial plasmids. Then, homologous recombination systems or PCR may be utilized in M. smegmatis or E. coli to perform gene replacement whereby the plasmid constructs containing the reporter genes ar put into mycobacteriophages.

A third method of introducing reporter genes and transcriptional promoters into mycobacteriophages is by use of transposons. For example, transposon IS1096 may be utilized. In order to use this methodology, reporter genes and transcriptional promoters are put into transposons, and the transposons containing the reporter genes and transcriptional promoters are delivered on plasmids in mycobacteria. Next, it is necessary to grow up the mycobacteriophages on a strain such as M. smegmatis, which strain contains the transposons. At certain frequencies, the transposons will hop into non-essential regions of the mycobacteriophages, thereby introducing themselves therein. The mycobacteriophages are still viable, and contain the reporter genes and transcriptional promoters.

A fourth method of introducing reporter genes and transcriptional promoters into mycobacteriophages is by debilitated phages packaged into phage heads and tails (phage particles). To utilize this methodology, it is necessary to develop helper phage systems which allow for pieces of DNA containing pac sites to be packaged. These helper phages allow for the synthesis of head and tail genes at will in mycobacteria, prevent themselves from being packaged into phage heads and tails, and facilitate packaging of pacmids into phag heads and tails. Helper phage systems may

be generated from the L5 mycobacteriophage. The genome of the helper phage is put into the mycobacterial chromosome, at which time the mycobacteria are grown up. Next, pacmids which comprise phages which have pac sites, reporter genes, transcriptional promoters and mycobacterial replicons are transformed onto the mycobacterial strain. The production of head and tail proteins may be induced, for example, through an increase in temperature, and the pacmids are then packaged into phage heads and tails. The L5 genome has cohesive (cos) termini. This suggests the possibility of constructing L5 cosmid vectors, which could be packaged through the cos sites into L5 particles either in vivo or in vitro. Then, a large number of genes could be easily and efficiently delivered to mycobacteria.

Packaging into phage heads and tails may also be utilized in a fifth methodology wherein the pacmid is a plasmid. The methodology is similar to the methodology wherein a debilitated phage is used, however, instead of using phage pacmids, the pacmids comprise plasmids which have pac sites, reporter genes, transcriptional promoters, and plasmid replicons.

Finally, direct cloning using recombinant DNA techniques in vitro may be used to introduce reporter genes and transcriptional promoters into

mycobacteriophages. This methodology consists of ligating a mycobacteriophage, identifying or introducing unique restriction enzyme sites in non-essential regions of the mycobacteriophage, cleaving the mycobacteriophage with the restriction enzyme sites, and cleaving DNA which encodes the promoter and the reporter gene so that it has the unique sites flanking it on either side. Next, ligation is set up in vitro between the cleaved mycobacteriophage with the unique restriction enzyme sites and the reporter gene cassette. The result is a circular DNA molecule which consists of the mycobacteriophage, the reporter genes and the transcriptional promoters. The circular DNA may then be electroporated directly into mycobacteria.

EXAMPLES

Expression of Reporter Gene lacZ and FFlux in Mycobacteria

A promoter probe vector was constructed which incorporated a truncated E. coli ß-galactosidase (lacZ) gene as a reporter probe into a shuttle plasmid vector that replicated in either mycobacteria or E. coli. Random DNA fragments from the three mycobacteriophages L1, TM4 and Bxb1 were cloned into a unique BamH1 site immediately upstream of the lacZ gene and screened for their ability to produce ß-galactosidase. This established that lacZ could be

used as a reporter gene in the mycobacteria, and identified the DNA sequences which could effectively express foreign genes in both M. smegmatis and M. tuberculosis. ß-galactosidase activity could be detected from lysed cells using OMPG, or from unlysed cells using either X-gal or a fluorescent methylumbelliferyl ß-galactosidase derivative. The promoter hsp60 gene highly expressed the lacZ gene in both M. smegmatis and BCG.

The FFlux gene was cloned into pMV261 downstream from the hsp60 promoter in plasmid pYUB180 (see Figure 2), which plasmid was shown to express the FFlux gene in M. smegmatis, BCG and M. tuberculosis H37Ra. The expression of the FFlux gene was detected by observing luminescence of mycobacterial clones containing the cloned gene in the dark room, and verified use in photographic film. This demonstrated that the luciferase was expressed in the mycobacteria, and that luciferin, the substrate used, was able to penetrate mycobacterial cell walls and yield photons expressed by the mycobacteria.

Detection of Photons In
Mycobacterial Cells Expressing FFlux

The expression of FFlux from the plasmid pYUB180 in M. smegmatis provided a model with which to determine a minimal number of individual cells

detectable with the luciferase assay. M. smegmatis containing pYUB180 were grown in the presence of kanamycin to ensure that every cell contained the plasmid. The cells were diluted 10-fold serially and the amount of luciferase activity was determined using a luminometer. Figure 3 shows that the amount of luciferase activity from 5×10^7 cells approached 10^8 luciferase units, though at this level of activity the luminometer was unable to yield an accurate measurement. However, the activity decreased in a linear manner down to 1200 units for 500 cells. Hence, 5000 cells expressing the FFlux gene can be clearly discerned above the background measurement, which approaches the number of cells that one would expect to observe in clinical samples.

Distinguishing Drug-Resistant Mycobacteria From Drug-Sensitive Mycobacteria Using Luciferase Activity

Since Firefly luciferase activity requires ATP, and ATP is produced only by living cells which are metabolically active, luciferase is a powerful indicator of the metabolic abilities of a bacterial cell. Since anti-tuberculosis drugs are likely to significantly decrease the metabolic activity of a cell, the measurement of luciferase activity should provide a sensitive means of distinguishing drug-resistant mycobacteria from drug-sensitive mycobacteria.

First, the kinetics of the production of luciferase activity of M. smegmatis containing pYUB180 following the addition of streptomycin, isoniazid, ethambutol, rifampicin, ciprofloxacin, novobiocin or cyanide, added at levels that inhibit the growth of M. smegmatis in plate assays, was measured.

As shown in Figure 4, Panel A, the levels of luciferase production were 100 to 1000 times less at eight hours after the addition of the drugs compared to the untreated control.

Next, this approach was used to distinguish drug-resistant from drug-sensitive mycobacteria. The pYUB180 deposit was transformed into streptomycin-resistant or novobiocin-resistant M. smegmatis mutants. Photon production by the drug-sensitive parent was compared to the streptomycin-resistant or novobiocin-resistant mutants. The drug-resistant mutants continued to produce luciferase activity levels comparable to the untreated patent in the presence of the appropriate antibiotic. In addition, the drug-resistant mutants produced 100 to 1000 times more luciferase activity than the drug-sensitive parent (see Figure 4, Panels B and C). Hence, a luciferase-based assay may be used to determine mycobacterial drug susceptibility.

Construction of TM4 Reporter Mycobacteriophages and Detection of Photons Following TM4::lux Infection

The first vectors developed to introduce recombinant DNA into mycobacteria were shuttle phasmid phage vectors. Shuttle phasmids have the ability to replicate in E. coli as cosmids and then replicate in mycobacteria as phages. Shuttle phasmids of TM4 which contained the FFlux and lacZ genes transcribed from hsp60 and L1 promoters, respectively, were constructed (see Figure 5).

A deposit of the shuttle phasmid (reporter mycobaceriophage) phAE39 which contains mycobacteriophage TM4, cosmid pYUB216, reporter gene FFlux and promoter hsp60, was made with the American Type Culture Collection on January 12, 1992 and catalogued as ATCC #75183. When the TM4::lux shuttle phasmid phAE39 was mixed with M. smegmatis cells, luciferase activity could be detected within 15 minutes of incubation, and continued to increase slightly over the next 4 hours (see Figure 6). These results show that the TM4::lux mycobacteriophage is capable of introducing the FFlux gene into mycobacterial cells, and that the FFlux gene can be expressed in mycobacteriophage-infected cells. Figure 7 represents a flow chart for cloning different promoters into the TM4::lux shuttle phasmid phAE39.

A deposit of the shuttle phasmid (reporter

mycobaceriophage) phAE37 which contains mycobacteriophage TM4, cosmid pYUB216, reporter gene lacZ and promoter L1, was made with the American Type Culture Collection on _____, 1992 and catalogued as ATCC #_____. The TM4::lacZ mycobacteriophage formed bright blue plaques when plated on media containing X-gal.

Construction of the L5 Reporter Mycobacterophage

Strategies for construction of the recombinant L5 mycobacteriophage may be investigated. The possibility of using the shuttle phasmid approach starting with L5 deletion derivatives, in which the size of the genome has been reduced, may also be explored. Initially, the largest gene 62 deletion available should be used. However, other deletion derivatives in which more of the gene 62-61-60 segment is lost should also be isolated. Another approach would be to attempt to introduce genes by homologous recombination with plasmids. Still another approach would be to transpose lux genes onto L5 using either the mini-Mu in vitro transposition system or a mycobacterial transposon such as IS1096.

Recombining reporter genes from recombinant plasmids onto L5 using a double recombination event may also be performed. This involves first constructing a recombinant plasmid that carries a reporter gen (lacZ may be more suitable) inserted

into gene 62 such that both the upstream and downstream parts of gene 62 are present. Advantages of this approach are that lacZ can be easily detected in agar media, that gene 62 is not an essential gene, and that lacZ is efficiently expressed from a promoter immediately upstream of gene 62. An L5 mycobacteriophage lysate may be prepared by growth of the plasmid-containing strain and recombinant mycobacteriophage progeny identified by plating the lysate on wild-type M. smegmatis for individual plaques on agar containing the indicator X-gal.

This recombination approach may be expanded to introduce other gene or DNA segments of the L5 genome. For example, it should be possible to add luciferase genes from FFlux in an identical manner, provided that packaging limits are not exceeded. In addition, inclusion of polylinker containing restriction enzyme sites unique for L5 would open the way for construction of L5 recombinants in vitro. Similar genetic strategies may be used to systematically reduce the size of the L5 genome by deletion of non-essential sequences.

Transposition offers an alternative method for the construction of reporter mycobacteriophages. A transposition system which is available is the mini-Mu in vitro transposition system. This is a defined biochemical reaction in which a mini-Mu transposon

carrying the desired gene is transposed onto the phage genome using purified MuA and MuB proteins. Similar transposition experiments have been tried with L5, but few L5 mini-Mu derivatives have been isolated. It is possible that this is due to the relatively large size of the transposon used. It is necessary to first construct a small Mu transposon which contains the reporter gene, a promoter and the two Mu in order for these experiments to be successful.

Development of L5 in vivo and in vitro Packaging Systems

g cosmids and packaging systems provide the efficiency of mycobacteriophage infection with the ability to inject large segments of non-mycobacteriophage DNA. Analogous mycobacterial systems would overcome packaging constraints encountered with recombinant mycobacteriophage genomes and allow the introduction of multiple copies or types of reporter genes into mycobacteria, potentially enhancing the sensitivity of the assay. In addition, they would help overcome any problems with host synthesis inhibition.

The development of L5 cosmids and packaging systems is dependent on the finding that the L5 genome contains cohesive termini. The g paradigm suggests that a relatively small region of DNA (approximately 500bp) around the cos site (in the ligated form) is

necessary to promote packaging. The first series of experiments with L5 would therefore be to identify the segment of the genome required for packaging by constructing a series of plasmids containing the L5 cos site and surrounding sequences. Cos activity may be determined by preparation of an L5 lysate on plasmid-containing M. smegmatis strains, followed by the identification of antibiotic-resistant transductants in the lysate, by transduction of M. smegmatis. This assay assumes that plasmid multimers of a total size of approximately 50kb are present in the cell and will be packaged. Although the presence of such multimers has not been demonstrated directly, they are likely to be generated by the homologous recombination system of M. smegmatis. If this assay should fail, cosmid vectors which contain both L5 g cos sites may be constructed. Insertion of 40-45kb of DNA (as in the construction of cosmid libraries) followed by g packaging in vitro and infection with E. coli will generate 50kb sized molecules containing L5 cos site. These should be isolated from E. coli and introduced by electroporation into M. smegmatis. Assuming that one of these approaches is successful, it would then be possible to define a small segment of L5 DNA required for packaging.

The construction of in vivo cosmid packaging

systems is a particularly attractive idea since it has proven very useful in E. coli. Thermoinducible lysogens of L5 may be suitable for in vivo packaging of L5 cosmids without further modification, since prophage excision may be a temperature-sensitive event. Efficient packaging of extrachromosomal cosmids present in the lysogen may be achieved by simple induction and growth at 42°C.

It is possible that some process other than excision is temperature-sensitive in lysogen induction. If so, it will be necessary to further debilitate the prophage in order to prevent DNA packaging of the prophage. There are a variety of ways to accomplish this. For example, the excise gene itself could be deleted (using a recombination strategy similar to that described above) such as to prevent excision. Another approach is to damage the cohesive termini (by exonucleolytic digestion) of an L5 thermoinducible derivative and construct a defective lysogen. A combination of approaches may be desirable, since even if prophage excision is a temperature-sensitive process, the destruction of cos might effectively reduce the background of spontaneous mycobacteriophage release.

Construction of in vitro packaging systems will follow similar lines. Extracts may be prepared from thermoinducible strains with non-packagable

prophages and assessed for their ability to package exogenously added L5 cosmid or mycobacteriophage DNA. Optimization of conditions should follow both empirical biochemical approaches and the well-established g systems. For example, it may be necessary to supplement the extracts with purified mycobacteriophage products such as the terminase or the tape-measure analogues (genes A/Nu and H of g respectively), neither of which have yet been identified.

Construction of Novel Shuttle Phasmids From Any Mycobacteriophage

Although mycobacteriophages L5 and TM4 can be used in the development of diagnostic luciferase and ß-galactosidase shuttle phasmids, there may be other mycobacteriophages, such as the mycobacteriophage DS6A which only infects BCG and M. tuberculosis strains, that might prove to have a more useful host range for clinical isolates. Diagnostic luciferase mycobacteriophages from these other mycobacteriophages may be developed by using the shuttle phasmid methodology described herein that has been proven successful for constructing mycobacteriophage vectors from both TM4 and phage L1.

Isolate Mycobacteriophage L5 and TM4 Mutants to Infect the Maximum Number of Clinial Isolates

For th diagnostic luciferase mycobacteriophage system to have maximal use in the clinical laboratory,

it will be essential that to develop a set of diagnostic mycobacteriophages that can efficiently infect any clinical isolate and possibly distinguish M. tuberculosis from M. avium and BCG. Both mycobacteriophages TM4 and L5 appear to have the ability to infect a large number of M. tuberculosis isolates. TM4 is very closely related to phage 33D, a mycobacteriophage that has been found not to infect every M. tuberculosis isolate used to define the mycobacteriophage typing schemes for M. tuberculosis isolates. However, this mycobacteriophage does not infect BCG. TM4 has been found to be almost identical by DNA hybridization and restriction analysis to 33D, and it shares the host-specificity with 33D in that it infects M. tuberculosis, but fails to infect BCG. mycobacteriophage L5 appears to share the same receptor as mycobacteriophage D29 which receptor has been previously shown to infect a very large number of M. tuberculosis isolates. L5, unlike 33D or TM4, infects all three morphotypes of M. avium including a wide range of serovariants.

If L5 or TM4 are found not to infect certain M. tuberculosis isolates, it may be possible to isolate mutants of these mycobacteriophages which plaque on the particular isolate. The inability to plaque on a particular isolate could result from the lack of a mycobacteriophage receptor or be the result

of lysogenization of the isolate with a homoimmune phage. Phage mutants with altered host range specificities or mutants which no longer bind a repressor (equivalent to virulent mutant of g) have been isolated in other systems. Variants of TM4 which can efficiently infect BCG have been isolated at frequencies of 10^7. Previous work has demonstrated that 33D, similarly to TM4, can not adsorb to BCG cells. Host-range variants of TM4 which not only plaque BCG, but also still plaque M. tuberculosis have been isolated. Similar strategies for M. tuberculosis isolates which are uninfected by L5 or TM4 may be used.

Detecting the Presence of
M. tuberculosis in Clinical Samples

The combined sensitivities of luciferase and mycobacteriophage infections should permit the detection of previously undetectable levels of M. tuberculosis cells in sputum, blood samples, or cerebral spinal fluid. A number of preliminary studies to optimize the detection of M. tuberculosis cells in a variety of body samples will be performed.

Detecting M. tuberculosis Grown In Primary
Human Macrophages and Macrophage Cell Lines

As a model system for optimizing detection of M. tuberculosis in infected monocytes and macrophages, primary human monocytes which have been purified by adherence for 1 hour or primary macrophages which have

been cultured for 6 days in microwells will be infected with M. tuberculosis H37Ra at varying multiplicities. The number of cells initially infected will be determined microscopically, and then at various periods of time from 2 hours to 30 days, the cells will by lysed by non-ionic detergent NP40 which has no effect on viability of mycobacteria, concentrated by centrifugation, plated for viable organisms and infected with the luciferase plasmids. Quantitative studies at different moi's and with varying numbers of infected cells will indicate how few bacilli/cell and bacilli/specimen can be detected.

The inability of M. tuberculosis cells isolated from macrophages to be infected with diagnostic shuttle phasmids could result from either the absence of the expression of the mycobacteriophage-receptor or the masking of the receptor with a membrane from a phagosome of the macrophage. The level of expression of phage receptors may be regulated by the environment in which the host cell is grown. For example, the g repressor of E. coli is induced by maltose and repressed by glucose. Studies to identify the receptors for mycobacteriophage L5 have been initiated. Similar studies for mycobacteriophage TM4 will also be performed. By identifying the genes encoding the receptor, it is possible to assay gene repression of the mycobacteriophage receptor of

M. tuberculosis cells when grown in macrophages by hybridization for the mRNA synthesis. If the receptor is not expressed in macrophages, it may be necessary to use a mycobacteriophage which recognizes a receptor that is constitutively expressed.

If the receptor is masked by a membrane of the macrophage, the cells isolated from macrophages may be treated with a variety of different detergents to find a treatment that would allow infection of the M. tuberculosis cells with the mycobacteriophages. Again, it may be necessary to cultivate the detergent-treated macrophages in broth for a few generations to gain expression of the receptors. The assays to determine the infectability of macrophages from mycobacteria include not only the luciferase assay for the TM4::lux mycobacteriophages, but also infectious centers assays in which free mycobacteriophages are removed and mycobacteriophage-producing cells are scored by a mixed plating on a lawn of M. smegmatis. This assay would be useful since infectability can be scored even if there are insufficient M. tuberculosis cells to form a bacterial lawn. It is important to re-evaluate the host range specificities of all of the mycobacteriophages in this assay. Free mycobacteriophages can simply be removed through the use of specific anti-mycobacteriophage antibodies.

Detecting M. tuberculosis in Sputum Samples

Sputum from a patient infected with M. tuberculosis contains a mixture of mucoploysaccharide, free M. tuberculosis cells, macrophages containing M. tuberculosis cells and a variety of cellular debris. Sputum samples from patients thought to have pulmonary tuberculosis may be used for a study in which various numbers of M. tuberculosis cells are added to sputum samples found to have no or few organisms by acid-fast staining. A variety of methods can be used to treat sputum samples so as to liquify the mucous and decontaminate the specimen under conditions in which all bacteria other than mycobacteria are killed. Because of the specificity of the phasmids, decontamination may not be as important as preserving the mycobacteriophage receptors. Nonetheless, the sputum samples may be treated initially with 2% w/v NaOH for 30 minutes at 37°C or with 0.5% N-acetyl cysteine + 1% NaOH. Alternatively, the sample may be treated with a variety of hydrolytic enzymes, such as collagenase, to help dissolve the sputum sample. If mycobacteriophage receptors are carbohydrates possibly sensitive to these conditions, other conditions may be utilized or the cells will be cultured 3-16 hours to allow recovery of infectivity before mycobacteriophage infection.

Detecting Mycobacteria In Blood Samples

Tuberculosis has been known to have a bacteremia. If the sensitivity necessary to detect 100 to 200 M. tuberculosis cells in a ml of sample can be obtained, levels of bacteremia in tuberculosis patients which were not previously observable may be observed. White cells should be purified over Ficoil-hypaque and lysed with 2% NP40, 1% SDS or freeze-thawing in the presence of DNAse to liberate intracellular mycobacteria. The pellet should then be infected with the diagnostic luciferase mycobacteriophage, or if only few organisms are present they can be concentrated by filtration onto filters, and filter areas cut out and infected.

Assuring Specificity On a Variety of Clinical Isolates and Species; Assessment of False Positives and Negatives

The luciferase assay may be optimized such that positive correlations of M. tuberculosis infections as indicated in the clinical lab may be obtained. The recombinant mycobacteriophages may be tested to ascertain the range of specificity that they have for other mycobacteria, and for the closely related genera Norcardia, Corynebacterium, and Actinomycetes strains. These strains may be obtained from the ATCC. A number of blinded tests including negative controls, M. tuberculosis-infected patients, samples

from patients infected with M. avium, and samples infected with other non-mycobacterial pathogens may be performed to ascertain the range of specificity.

The ability to rapidly assess the susceptibilities of M. tuberculosis isolates to isoniazid, ethambutol, rifampicin, pyrazinamide and other antibiotics will have a major impact on the treatment of tuberculosis patients. After the isolation of M. tuberculosis cells from a sputum sample, which may take several weeks, the assessment of drug-susceptibilities may take an additional 2 to 9 weeks. Diagnostic reporter mycobacteriophages may allow for evaluations of drug-susceptibilities at the time a sputum sample is collected. Alternatively, this approach would shorten the time necessary to assess drug-susceptibilities of purified M. tuberculosis colonies grown up from clinical samples.

Luciferase Assays for M. tuberculosis Cells in the Presence of Drugs

The results of the experiments suggest that by using luciferase as an indicator for the metabolic ability of the cell, it may be possible to define conditions which will enable us to distinguish drug-resistant mycobacteria from drug-sensitive mycobacteria. To test this hypothesis, isolated

mutants of M. tuberculosis H37Ra which are resistant to isoniazid, rifampicin, ethambutol, or pyrazinamide would be used to generate a set of cogenic mutants. These independent mutants and the parent strains would be transformed with pYUB180. Luciferase activity will be assessed in the presence and absence of drugs in order to determine the optimal conditions for distinguishing between drug-resistant and drug-sensitive cells. It is quite possible that the window of time to observe differences for different drugs could vary and require different incubation times for each drug.

The choice of the promoter for expressing luciferase may provide a needed parameter to more readily assess drug action. For example, in the case of E. coli, gyrase promoters are greatly stimulated in the presence of gyrase inhibitors.

Clinical isolates of M. tuberculosis may be transformed with PYUB180 and tested for luciferase activity in the presence and absence of drugs. The luciferase assays with mycobacteriophage infections with lux mycobacteriophages on in vitro-grown M. tuberculosis cells will first be optimized, and then extended to M. tuberculosis cells grown in macrophages or isolated from sputum samples.

Critical Assessment of Drug-Susceptibility Testing

As for the detection of M. tuberculosis from clinical samples, the luciferase assay may be optimized so that the drug-susceptibility patterns for any clinical isolate may be obtained. It may be possible to add diagnostic mycobacteriophages to a single clinical specimen, aliquot the mixture into various tubes and add antibiotic drugs. Thus every experiment would have an internal control and each drug-treated sample could be compared to an untreated control. The critical parameter to conclude drug-resistance or sensitivity lies in the comparison.

Although the invention herein has been described with reference to particular embodiments, it is to be understood that these embodiments are merely illustrative of various aspects of the invention. Thus, it is to be understood that numerous modifications may be made in the illustrative embodiments and other arrangements may be devised without departing from the spirit and scope of the invention.

WHAT IS CLAIMED IS:

1. A method of producing mycobacterial species-specific reporter mycobacteriophages which comprises introducing reporter genes and transcriptional promoters into the genomes of mycobacterial species-specific mycobacteriophages wherein upon incubation with the mycobacteria for which said reporter mycobacteriophage is specific, the reporter genes of said reporter mycobacteriophage will express a gene product which is detectable.

2. The method according to Claim 1 wherein the reporter genes and transcriptional promoters are introduced into the mycobacteriophages by shuttle phasmid technology.

3. The method according to Claim 1 wherein the reporter genes and transcriptional promoters are introduced into the mycobacteriophages by homologous recominbation or PCR.

4. The method according to Claim 1 wherein the reporter genes and transcriptional promoters are introduced into the mycobacteriophages by transposon technology.

5. The method according to Claim 1 wherein the r porter genes and transcriptional promoters are

introduced into the mycobacteriophages by debilitated phages packaged into page heads and tails.

6. The method according to Claim 1 wherein the reporter genes and transcriptional promoters are introduced into the mycobacteriophages by plasmids packaged into phage heads and tails.

7. The method according to Claim 1 wherein the reporter genes and transcriptional promoters are introduced into the mycobacteriophages by recombinant DNA techniques.

8. The method according to Claim 1 wherein the mycobacteria is M. tuberculosis.

9. The method according to Claim 1 wherein the mycobacterial species-specific mycobacteriophage is L5, TM4 or DS6A.

10. The method according to Claim 1 wherein the reporter genes are luciferase genes or the ß-galactosidase gene.

11. The method according to Claim 10 wherein the luciferase genes are selected from the group consisting of Firefly lux gene, Vibrio fischeri lux genes, Xenorhabdus luminescens lux genes and lacZ genes.

12. The method according to Claim 1 wherein the transcriptional promoter is hsp60 or the L5 gene 62 promoter.

13. The method according to Claim 1 wherein the gene product is photons.

14. The method according to Claim 1 wherein the gene product is made detectable by contacting said gene product with a substrate.

15. The method according to Claim 14 wherein the substrate is luciferin or decanal.

16. The mycobacterial species-specific reporter mycobacteriophage produced by the method of Claim 1.

17. A mycobacterial species-specific reporter mycobacteriophage comprising a mycobacterial species-specific mycobacteriophage which contains in its genome reporter genes and a transcriptional promoter, wherein the reporter genes express a gene product upon incubation with the mycobacteria for which the reporter mycobacteriophage is specific.

18. The mycobacterial species-specific reporter mycobacteriophage according to Claim 17 wherein the mycobacteria is M. tuberculosis.

19. The mycobacterial species-specific reporter mycobacteriophage according to Claim 17 wherein the mycobacterial species-specific mycobacteriophage is L5, TM4 or DS6A.

20. The mycobacterial species-specific r porter mycobacteriophage according to Claim 17

wherein the reporter genes are luciferase genes or the ß-galactosidase gene.

21. The mycobacterial species-specific reporter mycobacteriophage according to Claim 20 wherein the luciferase genes are selected from the group consisting of Firefly lux gene, Vibrio fischeri lux genes, Xenorhabdus luminescens lux genes and lacZ genes.

22. The mycobacterial species-specific reporter mycobacteriophage according to Claim 17 wherein the transcriptional promoter is hsp60 or the L5 gene 62 promoter.

23. The mycobacterial species-specific reporter mycobacteriophage according to Claim 17 wherein the gene product is photons.

24. The mycobacterial species-specific reporter mycobacteriophage according to Claim 17 wherein the gene product is made detectable by contacting said gene product with a substrate.

25. The mycobacterial species-specific reporter mycobacteriophage according to Claim 24 wherein the substrate is luciferin or decanal.

26. A method of diagnosing a mycobacterial disease which comprises incubating a sample which may contain myco- bacteria with mycobacterial species-specific mycobacteriophages which contain

reporter genes and transcriptional promoters in their genomes, wherein the reporter genes produce a gene product upon incubation with the mycobacteria for which the mycobacteriophage is specific, and wherein the gene product is detectable.

27. The method according to Claim 26 wherein the mycobacterial disease is tuberculosis.

28. The method according to Claim 26 wherein the mycobacteria is M. tuberculosis.

29. The method according to Claim 26 wherein the mycobacterial species-specific mycobacteriophage is L5, TM4 or DS6A.

30. The method according to Claim 26 wherein the reporter genes are luciferase genes or the ß-galactosidase gene.

31. The method according to Claim 30 wherein the luciferase genes are selected from the group consisting of Firefly lux gene, Vibrio fischeri lux genes, Xenorhabdus luminescens lux genes and lacZ genes.

32. The method according to Claim 26 wherein the transcriptional promoter is hsp60 or the L5 gene 62 promoter.

33. The method according to Claim 26 wherein the gene product is photons.

34. The method according to Claim 26 wherein

the gene product is made detectable by contacting said gene product with a substrate.

35. The method according to Claim 34 wherein the substrate is luciferin or decanal.

36. The method according to Claim 26 wherein the sample is blood or sputum.

37. A method of assessing drug resistance of a mycobacterial strain which comprises:

(a) incubating a sample which contains a myco- bacterial strain with mycobacterial species-specific mycobacteriophages which contain in their genomes transcriptional promoters and reporter genes which produce gene products;

(b) adding an anti-mycobacterial drug to the incubation; and

(c) detecting whether the gene product is present in the sample, such presence indicating drug resistance of the mycobacterial strain.

38. The method according to Claim 37 wherein the mycobacterial strain is a strain of M. tuberculosis.

39. The method according to Claim 37 wherein the mycobacterial species-specific mycobacteriophage is L5, or TM4 or DS6A.

40. The method according to Claim 37 wherein the reporter genes are luciferase genes or the ß-galactosidase.

41. The method according to Claim 40 wherein the luciferase genes are selected from the group consisting of Firefly lux, gene, Vibrio fischeri lux genes, Xenorhabdus luminescens lux genes and lacZ genes.

42. The method according to Claim 37 wherein the gene product is photons.

43. The method according to Claim 37 wherein the transcriptional promoter is hsp60 or the L5 gene 62 promoter.

44. The method according to Claim 37 wherein the anti-mycobacterial drug is selected from the group consisting of streptomycin, isoniazid, ethambutol, rifampicin, ciproflo- xacin, novobiocin and cyanide.

45. The method according to Claim 37 wherein the gene product is made detectable by contacting said gene product with a substrate.

46. The method according to Claim 45 wherein the substrate is luciferin or decanal.

47. The method according to Claim 37 wherein the sample is blood or sputum.

INTERNATIONAL SEARCH REPORT

Inte...ational application No.
PCT/US93/00913

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C12N 07/00; C12P 21/06; C12Q 01/66
US CL :435/ 235.1, 69.8, 8
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/ 235.1, 69.8, 8

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

—

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

medline, dialog, aps

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US, A, 4,861,709 (Ulitzer et al.) 29 August 1989, see column 8, 14-17, and 28.	1-47
Y	Revista Cubana de Medicina Tropical, Volume 41, No. 2, issued 1989, M.C.A. Jimenez et al., "Phage Typing Marker Study of Mycobacterium-Tuberculosis Strains from Ethiopia Preliminary Report", pages 192-199, see abstract.	1-47
Y	Nature, Volume 351, issued 06 June 1991, C.K. Stover et al., "New Use of BCG for Recombinant Vaccines", pages 456-460, see entire document.	1-47

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date		
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search 15 APRIL 1993	Date of mailing of the international search report 21 APR 1993
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. NOT APPLICABLE	Authorized officer KATHLEEN L. CHOI Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/00913

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Folia Microbiology, Volume 36, No. 5, issued 1991, J. Konicek et al., "Gene Manipulation in Mycobacteria", pages 411-422, see pages 415-417.	1-47
Y	Nature, Volume 327, issued 11 June 1987, W.R. Jacobs Jr. et al., "Introduction to Foreign DNA into Mycobacteria using a Shuttle Phasmid", pages 532-535, see entire document.	1-47
Y	Proceedings of the National Academy of Sciences, Volume 88, issued April 1991, M.H. Lee et al., "Site-specific Integration of Mycobacteriophage L5: Integration-proficient Vectors for Mycobacterium smegmatis, Mycobacterium tuberculosis, and bacille Calmette-Guerin", pages 3111-3115, see entire document.	1-47
Y	Journal of General Virology, Volume 26, No. 1, issued January 1975, J.A. Hewitt, "Miniphage - a Class of Satellite Phage to M13", pages 87-94, see abstract.	5, 6
Y	Journal of Bacteriology, Volume 149, No. 3, issued March 1982, M.J. Orbach et al., "Transfer of Chimeric Plasmids among Salmonella typhimurium Strains by P22 Transduction", pages 985-994, see entire document.	5, 6
Y	Zentralbl. Veterinaermed., Reihe B, Volume 25, No. 5, issued 1982, R. Weiss et al., "Resistance Testing of Bacteria by Firefly Bioluminescence. A Rapid Test", pages 359-71, see abstract.	37-47
Y	Fortschr. Veterinaermed., Volume 35, issued 1982, R. Weiss et al., "Bioluminescent Methods to Test the Antibiotic Sensitivity of Bacteria", pages 323-328, see abstract.	37-47